

## Reviewed Preprint

v1 • February 27, 2026

Not revised

## Reviewed Preprint

v2 • April 21, 2026

Revised by authors

## ✉ For correspondence:

[yoram.bonneh@gmail.com](mailto:yoram.bonneh@gmail.com)

## Competing interests: No

competing interests declared

## Reviewing editor: Krystal R Huxlin,

University of Rochester, United

States

© 2026, Pinchuk-Yacobi et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

# Serial Dependence Predicts Generalization in Perceptual Learning

Noga Pinchuk-Yacobi<sup>1,2</sup>, Dov Sagi<sup>3</sup>, Yoram S Bonneh<sup>1,2</sup> ✉

<sup>1</sup>School of Optometry and Vision Science, Faculty of Life Science, Bar-Ilan University, Ramat-Gan, Israel • <sup>2</sup>The Leslie and Susan Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel • <sup>3</sup>Department of Brain Sciences, The Weizmann Institute of Science, Rehovot, Israel

## eLife Assessment

This **fundamental** study describes long-range serial dependence of performance on a visual texture discrimination training task that manipulated conditions to induce differing degrees of location transfer of learning. The authors re-analyzed previously-published, behavioral data, generating **compelling** evidence from converging approaches that the serial dependence effects persist over multiple days of training, and may share a common causal mechanism with training-induced location transfer. By informing our understanding of the importance of temporal integration to long-term perceptual learning and its propensity towards specificity or generalizability, these results should interest neuroscientists who seek to uncover underlying neural mechanisms for these processes.

<https://doi.org/10.7554/eLife.109830.2.sa3>

## Abstract

Visual perception is shaped by recent experience, but how these momentary influences accumulate to support long-term learning and generalization remains unclear. Here, we asked whether short-term memory traces, attractive serial-dependence effects (SDEs), promote learning generalization. We re-analyzed over 200,000 trials from observers trained on a visual texture-discrimination task under three conditions that differentially modulated generalization. Under certain conditions, SDEs reached further back in time than previously reported and persisted after eight days of practice, despite the non-informative nature of past stimuli. Observers in conditions that promoted generalization displayed larger long-range SDEs, and individual SDE magnitude predicted transfer of learning across locations. We propose that SDE is associated with learning flexibility, providing a principled framework for when and why perceptual learning generalizes, which is central to theories of cognitive flexibility. Attractive serial dependence is not an extra mechanism in this model—it is the behavioral footprint of ongoing template plasticity required for flexibility in changing environments.

## Introduction

Perceptual learning (PL) and serial dependence effects (SDEs) are two fundamental processes that shape how we perceive and interpret sensory information. Although both rely on perceptual memory traces, they operate over distinct timescales and have been traditionally studied separately. PL leads to long-lasting improvements in sensory discrimination following repeated training (Sagi, 2011 [↗](#)). However, in many cases, improvements remain limited to the trained stimulus or location, while in others, learning generalizes to new contexts. What determines whether learning stays local or generalizes is still not fully understood (Cheng et al., 2025 [↗](#); Lu & Doshier, 2022 [↗](#)).

SDEs reflect short-term biases that pull current perceptual judgments toward what was recently seen or chosen (Abrahamyan et al., 2016 [↗](#); Braun et al., 2018 [↗](#); Falmagne et al., 1975 [↗](#); Fischer & Whitney, 2014 [↗](#); Fornaciai & Park, 2018 [↗](#); Liberman et al., 2014 [↗](#); Suárez-Pinilla et al., 2018 [↗](#); Urai et al., 2019 [↗](#)). These biases are often interpreted as reflecting a Bayesian integration process that combines prior information with current input to enhance perceptual stability and efficiency, particularly under uncertainty, based on the assumption that the natural environment is typically stable (Cicchini et al., 2018 [↗](#); Fischer & Whitney, 2014 [↗](#); Manassi & Whitney, 2024 [↗](#)). However, in tasks where stimuli vary randomly across trials, such biases can actually impair performance, yet they persist. This suggests they may serve an additional role beyond what was previously proposed. One possibility is that they reflect ongoing trial-by-trial updates to internal decision templates that persist over time and may bridge short-term memory and long-term learning (Fritsche et al., 2017 [↗](#); Pascucci et al., 2019 [↗](#)).

In this work, we examine how serial dependence interacts with perceptual learning using the texture discrimination task (TDT; Karni & Sagi, 1991 [↗](#)). The role of serial dependence in the TDT is likely complex, potentially influencing both immediate perceptual judgments and long-term learning dynamics. In the short term, biases toward the orientation of prior targets, randomly varying in typical experimental setups, and thus irrelevant to the current trial, may impair performance by introducing decision noise. At the same time, temporal integration of stimulus representations across trials may help reduce uncertainty and extract structure from noisy input. Normative analyses predict that the benefit of engaging costlier, memory-dependent integration should peak at intermediate uncertainty, while being limited in highly certain conditions (Tavoni et al., 2022 [↗](#)). Over extended training, serial dependence may evolve within and across daily sessions as learning progresses. Serial dependence may also arise as a result of trial by trial network updates during learning (Petrov et al., 2006 [↗](#), 2005 [↗](#)), depending on how the effect of recent stimuli is weighted during network updates. Specifically, we ask whether serial dependence can serve as a marker of short-term temporal integration that contributes to long-term generalization of learning. This idea builds on the hypothesis that broader temporal integration, covering a larger network state space, may reduce overfitting to local stimulus features and allow learning to generalize more broadly.

To test this, we reanalyzed data from a large-scale perceptual learning study in which observers practiced the TDT under three training conditions designed to modulate the degree of learning generalization (Harris et al., 2012 [↗](#)). The study showed that learning generalized to a new, untrained location when targets appeared randomly across two locations or were intermixed with target-less (dummy) trials. In contrast, learning became location-specific when targets consistently appeared in a single location in all trials, likely due to increased sensory adaptation (Karni & Sagi, 1991 [↗](#)). These findings were attributed to differences in adaptation state: unadapted networks supported spatial transfer, whereas adaptation induced localized plasticity that constrained generalization. However, the memory mechanisms underlying this flexibility remain unclear. By examining serial dependence within this paradigm, we aim to gain further insight into these mechanisms and better understand how recent visual experiences influence both immediate perceptual reports and long-term learning outcomes. Specifically, we ask whether training conditions that promote generalization are associated with stronger or longer-lasting SDEs, suggesting a new role for serial dependence in contributing to broader learning transfer. We consider a learning mechanism where adaptation dependent inhibition controls decision template updates, and by that learning flexibility. On this account, serial dependence is a consequence of network plasticity.

## Methods

We reanalyzed data from 50 observers who participated in the texture discrimination task (TDT) as described by Harris, Gliksberg & Sagi (2012) [↗](#). In this dual-task experiment, observers identified the orientation (vertical or horizontal) of a target composed of three peripheral diagonal lines embedded within a uniform background of horizontal lines while simultaneously performing a forced-choice letter discrimination task (T vs L at the center of the stimulus) to

maintain fixation (Figure 1A). The experiment consisted of four daily sessions with the target presented at one location (or two fixed locations in the 2loc condition), followed by four additional daily sessions at a second location (or a second pair of fixed locations in the 2loc condition). Performance on the TDT task, the TDT threshold, is quantified as the SOA (see Figure 1) that yields 78% correct discrimination ( $SOA_{\text{threshold}}$ ). The reaction time (RT) used in the analysis was defined as  $RT(\text{TDT}) - RT(\text{fixation task})$ , where RT for each task was measured from stimulus onset.

Participants were assigned to one of three experimental conditions:

- **1loc condition:** The target consistently appeared at the same fixed location across all trials.
- **2loc condition:** The target randomly appeared at one of two diagonally opposite locations with equal eccentricity, alternating between trials.
- **Dummy condition:** The target appeared at a fixed location, but genuine target trials were randomly interleaved with ‘dummy’ trials, where no target was present (replaced by background elements).

For detailed specifications of the groups assigned to each condition, see Table S1.

We quantified serial dependence effects (SDE) to examine how prior visual experiences influence current perceptual reports (Figure 1B), and whether these dependencies affect learning specificity across experimental conditions. To estimate SDE, we fit a linear mixed-effects (LME) model (Figure 1C) that evaluated the influence of prior target stimuli (up to 10 trials back) on current reports. For each lag  $n$ , the model estimates a coefficient  $W_n$ ; we refer to this coefficient simply as  $W$ -bias. These coefficients represent the magnitude of the serial-dependence effect for each  $n$ -back stimulus. We modeled the report probability as

$$P(T_0 | T_0) = P_0 + \sum_{n=1}^{10} W_n \cdot sT_n$$

Where:

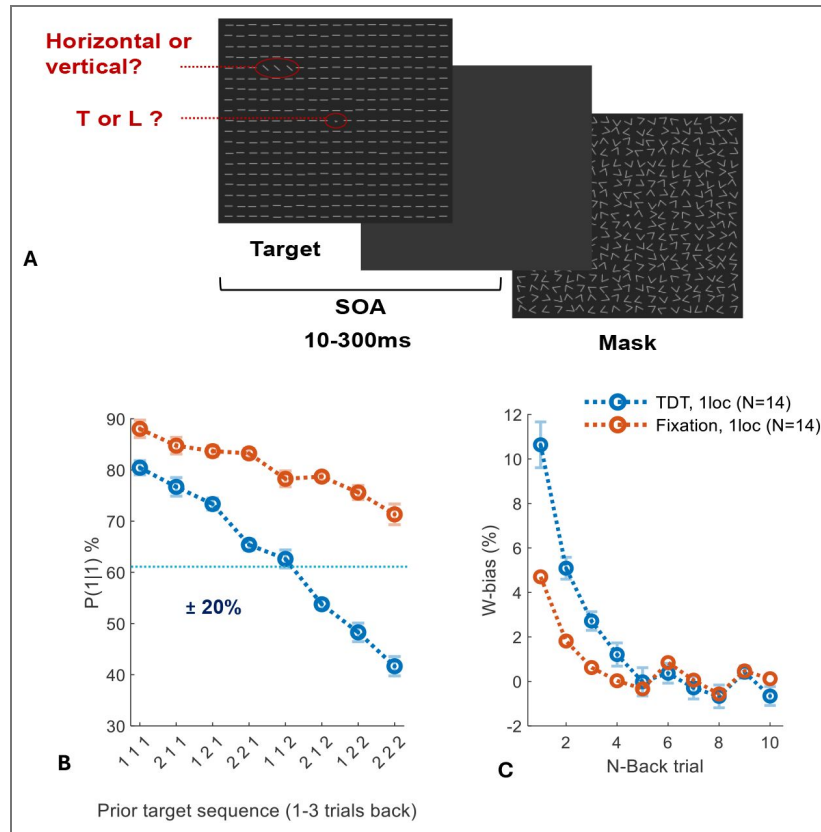
- $P(T_0 | T_0)$ : probability of reporting target  $T_0 \in \{+1, -1\}$  when the target is  $T_0$ .
- $P_0$ : history-independent baseline report probability.
- $sT_n$ : +1 if the  $n$ -back target’s orientation matches the current target (e.g., both vertical or both horizontal), -1 otherwise.
- $W_n$ ( $W$ -bias): change in report probability (bias) associated with the  $n$ -back target.

Fixed effects accounted for the orientations of these stimuli, while a random intercept captured individual differences across observers. Interaction terms were excluded, as statistically significant interactions were small and inconsistent. To preserve the precise temporal structure of the data, all trials were included in the sequential  $n$ -back count across all experimental conditions. In the Linear Mixed Effects (LME) analysis, we modeled these trial types using distinct regressors: each  $n$ -back lag included separate predictors for visible and invisible targets, further differentiated by trial type (dummy vs. target) and relative location (ipsilateral vs. contralateral) where applicable. The SDE values reported here reflect only the influence of relevant target-present history trials; the effects of other history types (e.g., dummy trials), while estimated to ensure the temporal integrity of the model, are not presented.

To systematically quantify serial dependence across different temporal scales, we defined three summary measures:

- **SDE-all:** The cumulative bias from the 10 preceding trials (1–10 back), capturing the total influence of recent history on current perception.
- **SDE-recent:** The cumulative bias from trials 1–3 back, reflecting the effect of very recent stimuli.
- **SDE-distant:** The cumulative bias from trials 4–6 back, representing the influence of more distant past trials.

All measures are expressed as percentage changes in report probability. For most analyses and figures, we filtered trials to include those that produced stronger SDEs. The specific filtering applied is motivated by technical issues. When using percent correct as a measure of performance, bias cannot be reliably estimated at or near ceiling performance, as correct responses leave little room for bias to manifest. However, easy trials can be taken as reliable references to condition



**Figure 1. Texture Discrimination Task (TDT) and Serial Dependence Effect (SDE)**

(A) TDT Trial Sequence: Observers identified the orientation (vertical or horizontal) of a target composed of three diagonal lines embedded within a background of horizontal lines while simultaneously performing a forced-choice letter discrimination task (T vs L) to maintain fixation. The 10 ms target frame was followed by a 100 ms patterned mask, with the stimulus onset asynchrony (SOA) between target and mask (10 to 300 ms) randomized across trials. (B) SDE via history sequence: Correct identifications ('Hits') increased by approximately 20% when the current target orientation matched the preceding three orientations (e.g., '1' preceded by '111') and decreased by about 20% with mismatching orientations (e.g., '1' preceded by '222'), relative to average performance (dotted light blue line). '1' and '2' represent the two possible orientations, either vertical and horizontal, or vice versa. The fixation task (red) showed much smaller biases, likely due to its higher overall performance. (C) SDE via Linear Mixed Effects (LME) Weights (W-bias, %): Influence of 1–10 back trials on current report. Summing the W-bias values (%) from the 1st, 2nd, and 3rd prior trials corresponds to the  $\pm 20\%$  bias for 1–3 back trials shown in panel B. Panels B and C include data from the 1loc condition (N=14) pooled across all training days (Days 1–8; 4 days at the first location and 4 days at the second location to assess generalization), including only trials with low-visibility current targets ( $SOA < SOA_{\text{threshold}} + 20$  ms, calculated on a per-subject basis). Blue represents the texture discrimination task, and red indicates the letter discrimination (fixation control) task.

the bias on, since low performance introduces uncertainty as for the perceptual effect of the reference orientation. Specifically, trials in which current targets were barely visible ( $SOA < SOA_{\text{threshold}} + 20$  ms), prior targets were highly visible ( $SOA > SOA_{\text{threshold}}$ ), and both appeared at the same location across trials (Figure 2). The exception was the SDE derived from the history sequence analysis (Figure 1B), where prior-target visibility could not be filtered. Therefore, when comparing this with the LME-based estimate (Figure 1C), only the current-target visibility filter was applied. Finally, to verify that our findings are not limited to these filtering choices, we also conducted control analyses including all prior-trial history regardless of visibility; these results are presented in Supplementary Figure S3 and confirm the robustness of our main findings.

For individual-level analyses assessing the relationship between SDE strength and TDT learning and generalization, we accounted for individual differences in  $SOA_{\text{threshold}}$  by equalizing the number of highly visible prior trials and restricting trials to SOAs between  $SOA_{\text{threshold}}$  and  $SOA_{\text{threshold}} + 140$  ms.

Because filtering on highly visible prior targets means observers usually reported them correctly, the key-history sequence closely tracked the target-history sequence, producing highly similar SDEs (Figure 2A, high-vis history). Accordingly, we present only Target-history results and omit Key-history plots for brevity.

For computing correlations on individual participants (Figures 4, 7), we used orthogonal regression without excluding outliers, then, we used the Pearson correlation to compute R and p-value. Confidence sleeves around regression lines were estimated using nonparametric bootstrap resampling (1,000 iterations). For each bootstrap sample, an orthogonal regression line was fit, and the 95% confidence sleeve was defined by the 2.5th and 97.5th percentiles of the predicted values across resamples. For non-significant correlations, we omitted the orthogonal regression line but kept the confidence sleeves.

**Data Availability.** The data that support the findings of this study are available from the corresponding author upon request.

**Ethics.** This manuscript presents a secondary analysis of previously published human behavioural data from Harris, Gliksberg & Sagi (2012). Original procedures received ethics approval, were conducted in accordance with the Declaration of Helsinki, and all participants provided written informed consent.

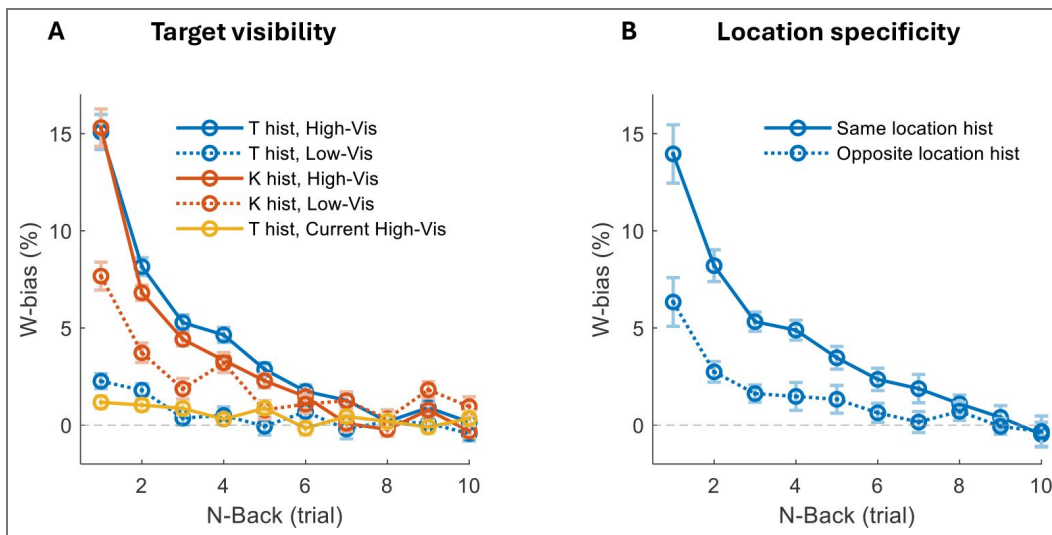
## Results

### Serial Dependence Effects (SDEs)

Observers' reports showed a significant 15% bias toward the orientation of the immediately preceding target (1-back; W1), indicating serial dependence. When the influence of the full range of past trials was considered (SDE-all), the bias increased to about 40% (Figure 2A). These values were measured under filtering conditions that enhanced the expression of serial dependence (detailed in the next section, *Conditions enhancing SDE: influence of target visibility and location*). Importantly, these biases were not attributable to motor responses, as they persisted in a dual-task setup and were modulated by the spatial location and visibility of target stimuli, independent of motor actions (Figure 2).

### Conditions enhancing SDE: influence of target visibility and location

Biases were most pronounced under conditions where the current targets were barely visible ( $SOA < SOA_{\text{threshold}} + 20$  ms) and prior targets were clearly visible ( $SOA > SOA_{\text{threshold}}$ ) with SDE-all reaching  $40 \pm 3\%$  (Figure 2A; Supplementary Figure S2A). Biases were significantly reduced when the current targets were highly visible (SDE-all =  $5 \pm 1\%$ ; reduction:  $35 \pm 3\%$ ;  $t(49) = 11.9$ ,  $p < 0.0001$ , Cohen's  $d = 1.7$ ) or when prior targets had low visibility (SDE-all =  $5 \pm 1\%$ ; reduction:  $35 \pm 3\%$ ;  $t(49) = 13.5$ ,  $p < 0.0001$ , Cohen's  $d = 1.9$ ). In the 2loc condition, biases were predominantly



**Figure 2. Serial dependence effects across trial history**

(A) Target visibility: SDEs (W-bias) were strongest when current targets had low visibility (high uncertainty) and prior targets had high visibility. Blue lines show target (T) histories and red lines show key (K) histories, with solid vs. dotted lines indicating prior high-vs. low-visibility trials. Visibility had a much weaker effect on key histories compared to target histories, with key-driven SDEs remaining high even when prior targets were invisible. The yellow line shows target history with high visibility in both history and current trials, demonstrating that SDEs are strongly reduced when current targets are highly visible (certainty) ( $N = 50$ ). (B) Location specificity: SDEs were larger when history originated from the same location as the current target compared to a diagonally opposite location (2loc condition,  $N = 14$ ). Error bars represent the standard error of the mean. The horizontal dashed line indicates zero bias.

location-selective, with significantly stronger effects when trial history originated from the same location (SDE-all =  $41 \pm 3\%$ ) compared to a diagonally opposite location (SDE-all =  $15 \pm 3\%$ ; reduction:  $26 \pm 3\%$ ;  $t(13) = 7.8$ ,  $p < 0.0001$ , Cohen's  $d = 2.1$ ; [Figure 2B](#); [Supplementary Figure S2B](#)).

## Decay of SDE over trials and with longer RT

Biases gradually decayed across successive trials but remained substantial, extending far into trial history ([Figure 2](#) and [Figure 6A](#)). In the 1loc condition, biases were significant up to 4 trials back ( $p < 0.001$  for  $N \leq 4$ ). In the 2loc condition, biases persisted longer, remaining significant up to 8 trials back ( $p < 0.001$  for  $N \leq 7$ ;  $p < 0.05$  for  $N = 8$ ). For contra 2loc condition, biases were significant up to 5 trials back ( $p < 0.001$  for  $N \leq 2$ ,  $p < 0.01$  for  $N = 3, 4$ ;  $p < 0.05$  for  $N=5$ ). The dummy condition showed the most prolonged biases, with significant effects extending up to 9 trials back ( $p < 0.001$  for  $N \leq 6$ ;  $p \leq 0.01$  for  $N = 7, 9$ ), although the effect at 8-back was not significant ( $p = 0.77$ ) and 10-back was borderline significant ( $p = 0.05$ ). Our analysis focused on decay across trials rather than elapsed time, as doubling inter-trial interval had no impact on the short-history effects but only attenuated the long-history effects (as observed when comparing the two groups in the 1loc condition; see [Table S1](#) for group details).

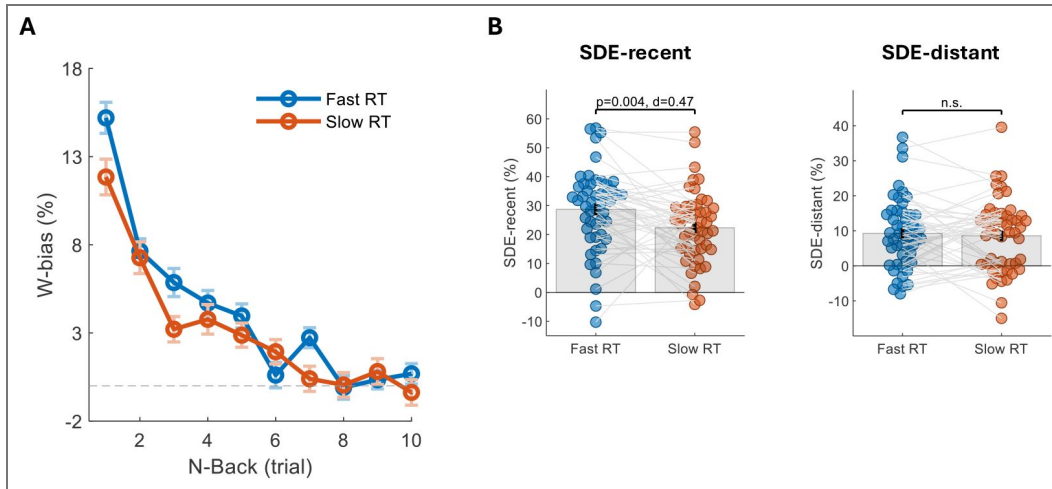
Motivated by recently found mechanism-dependent relationship between response bias and reaction time (RT) ([Dekel & Sagi, 2020](#)), we examined the dependence of SDE on RT. We compared SDEs in trials with the fastest and slowest RTs. For each observer, trials were divided into quartiles based on RTs calculated separately for each training day, to account for overall reductions in RT with practice. SDEs were then computed using all trials from the fastest quartile (lowest 25%) and the slowest quartile (highest 25%) ([Figure 3A](#)). Recent SDEs were significantly stronger for fast RT (SDE-recent =  $29 \pm 2\%$ ) compared to slow RT (SDE-recent =  $22 \pm 2\%$ ), yielding a reduction of  $7 \pm 2\%$  ( $t(49) = 3$ ,  $p < 0.01$ , Cohen's  $d = 0.4$ ; [Figure 3B](#), left panel; [Supplementary Figure S2C](#)), whereas distant SDEs did not significantly differ between RT conditions (SDE-distant =  $9 \pm 1\%$  for both fast and slow RT;  $t(49) = 0.5$ ,  $p = 0.64$ ; [Figure 3B](#), right panel; [Supplementary Figure S2C](#)). This difference between recent and distant SDEs suggests that they may arise from distinct underlying mechanisms. Observers from all conditions were combined in this analysis to increase statistical power, as separate analyses revealed a similar qualitative pattern across conditions: recent SDEs were stronger for fast compared to slow RTs, reaching significance in the 1loc and 2loc conditions but not in the dummy condition, likely due to smaller sample size. In contrast, distant SDEs showed no significant RT-related change in any condition.

## Dynamics of SDE across days and locations

Despite randomized target orientations across trials, rendering past orientations irrelevant to current judgments, serial dependence biases remained strong and highly significant across all training days ( $p < 0.0001$  for each day; [Figure 4A](#)). A two-factor repeated-measures ANOVA with Location (first vs. second retinotopic site) and Day-in-block (1–4) as within-subject factors revealed a significant main effect of Day-in-block,  $F(3, 147) = 4.7$ ,  $p < 0.01$ , indicating modest decrease of SDE magnitude across training days. The main effect of Location was not significant ( $F(1, 49) = 0.08$ ,  $p = 0.78$ ), with comparable SDE-all values at the first trained location (Days 1–4:  $40 \pm 3\%$ ) and the second trained location (Days 5–8:  $41 \pm 3\%$ ). There was no Location  $\times$  Day-in-block interaction ( $F(3, 147) = 0.47$ ,  $p = 0.71$ ), indicating similar temporal dynamics across locations. Post-hoc comparisons (Bonferroni-corrected) revealed a modest but statistically reliable decrease in SDE from Day 7 to Day 8 (the last two sessions at the second location;  $p < 0.01$ ), whereas no other day-to-day differences reached significance. No significant correlation was found between biases and SOA thresholds across observers ( $r = -0.13$ ,  $p = 0.37$ , average across days 1-8), nor between biases and improvements in performance at the first location ( $r = -0.09$ ,  $p = 0.54$ , average across days 1-4), suggesting that the magnitude of serial dependence does not predict the overall amount of perceptual learning ([Supplementary Figure S1](#)).

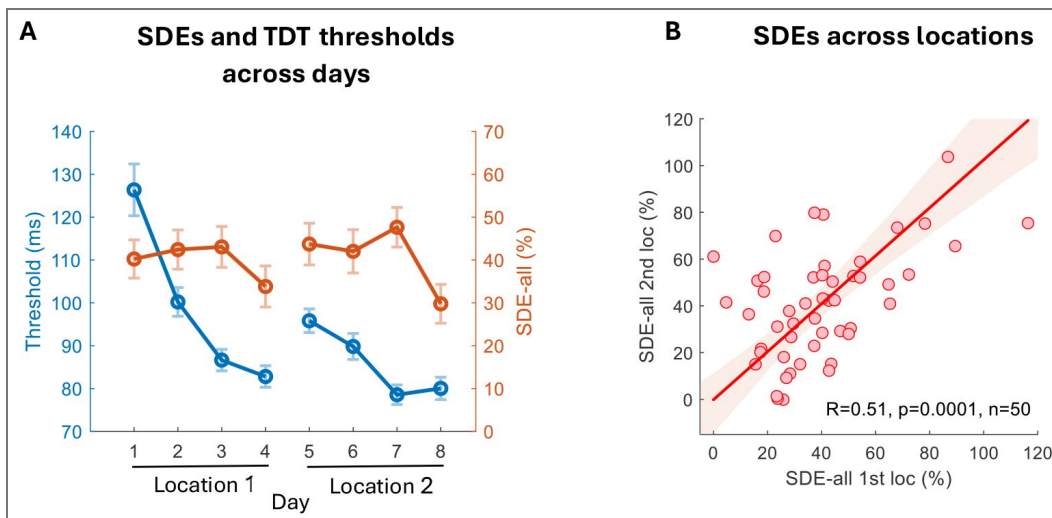
**Figure 3. Effect of reaction time on SDEs.**

(A) SDEs (W-bias) as a function of N-back trial, calculated separately for the fastest (first 25%) and slowest (last 25%) RT quartiles, defined per day within each observer. SDEs were then computed on the corresponding fast and slow trials and averaged across observers. Biases were stronger for fast RTs, particularly at recent lags. (B) Paired comparisons of recent and distant SDEs for fast vs. slow RTs. Recent SDEs were significantly higher for fast RTs (left panel), whereas distant SDEs did not differ between RT conditions (right panel). Gray bars indicate group means; dots and connecting lines represent individual observers ( $N = 50$ ).



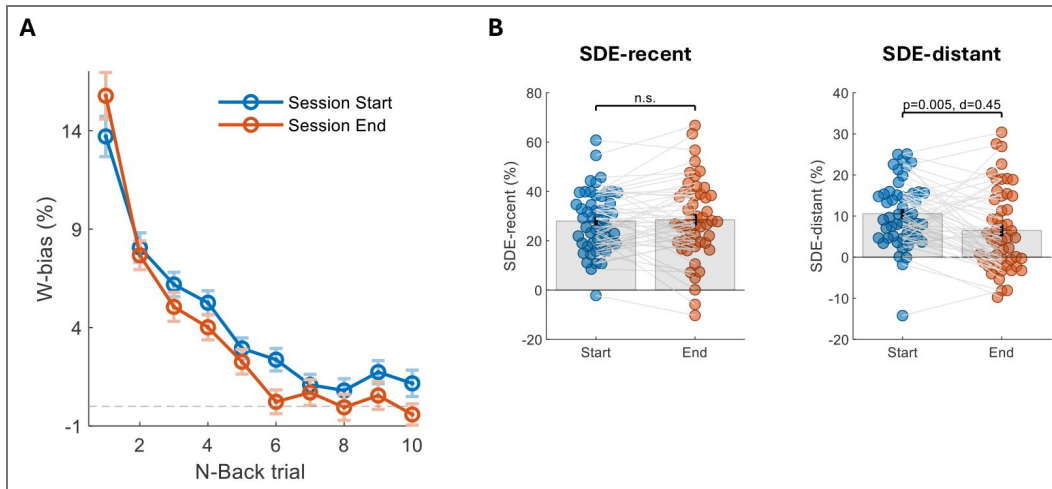
**Figure 4. Dynamics of SDE across days and locations**

(A) SDEs and TDT thresholds across days: Serial dependence (SDE-all, red) remained strong and consistent throughout the eight training days, despite large improvements in TDT thresholds (blue). A small reduction in SDE was observed across days and reached significance only between Days 7–8. The target location was changed after Day 4. (B) SDEs across locations: Correlation of SDE-all between the first (Days 1–4) and second (Days 5–8) trained locations across observers ( $N = 50$ ). The strong correlation indicates that the magnitude of serial dependence is a stable observer-specific trait, consistent across retinotopic locations. In (B), shaded regions denote 95% bootstrap confidence sleeves around the orthogonal regression fit.



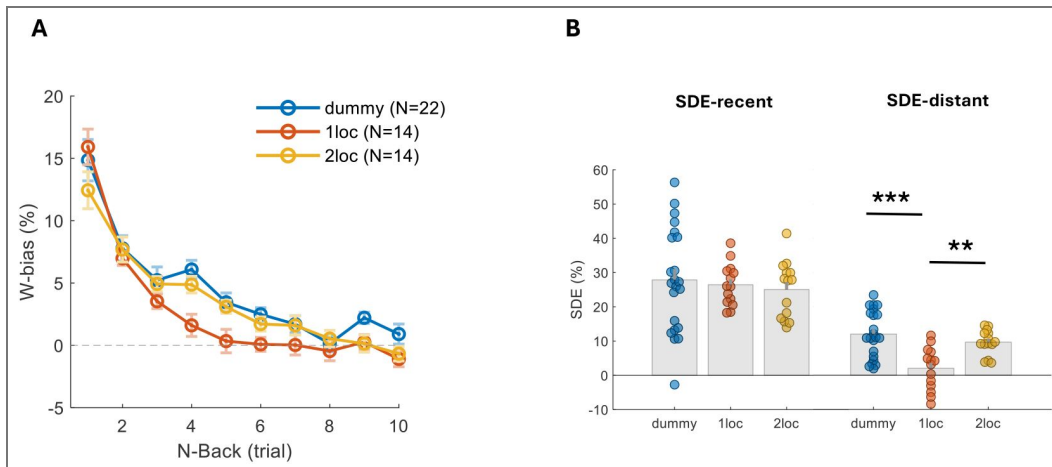
**Figure 5. Within-session dynamics of SDE**

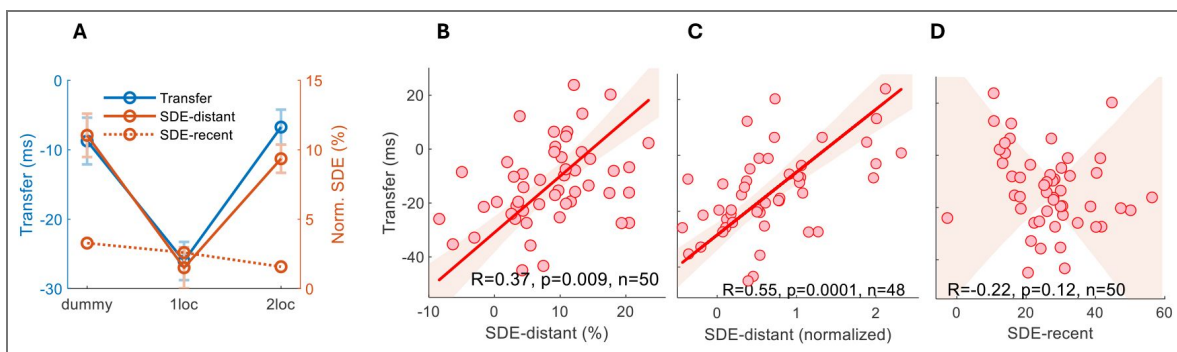
(A) SDEs (W-bias) as a function of N-back trial, computed separately for the first (blue) and last (red) third of each session. Biases showed an overall decrease across the session, while the 1-back bias (W1) increased slightly. (B) Recent and distant SDE components. Recent SDEs remained stable across the session (left panel), whereas distant SDEs showed a significant reduction (right panel). This pattern is consistent with sensory adaptation developing over the course of the session, selectively attenuating serial dependence from more temporally distant trials ( $N = 50$ ).



**Figure 6. SDEs across experimental conditions and trial history**

(A) Mean SDEs (W-bias) as a function of trial history (N-back) for the three experimental conditions: dummy (blue,  $N = 22$ ), 1loc (red,  $N = 14$ ), and 2loc (yellow,  $N = 14$ ). In the 1loc condition, biases decayed more rapidly, while in the 2loc and dummy conditions they persisted further back in trial history. (B) SDEs across individual observers ( $N = 50$ ), shown separately for recent lags (1–3 back; left panel) and distant lags (4–6 back; right panel). Each dot represents one observer; gray bars indicate group means. Recent SDEs were consistent across conditions, whereas distant SDEs were significantly stronger in the 2loc and dummy conditions compared to the 1loc condition ( $***p \leq 0.001$ ,  $**p \leq 0.01$ ).





**Figure 7. Relationship between SDE and learning transfer**

(A) Group-level comparison of learning transfer (blue), SDE-distant (solid red), and SDE-recent (dotted red) across the three experimental conditions (dummy, 1loc, 2loc). Transfer and SDE values are plotted on separate axes, with SDE measures normalized by subtracting the mean of the 1loc condition. Conditions showing greater learning generalization (dummy, 2loc) also exhibited stronger SDE-distant effects. In contrast, SDE-recent was relatively constant across conditions, suggesting that generalization was primarily linked to distant serial dependence. (B) Across observers, learning transfer correlated positively with SDE-distant ( $r = 0.37, p < 0.01, N = 50$ ), indicating that stronger distant serial dependence predicted greater generalization. (C) SDE-distant values were normalized to the 1-back effect to estimate the temporal decay constant of SDE, reflecting how long biases persisted across trials. Observers with longer decay constants showed greater learning transfer ( $r = 0.50, p < 0.001, N = 48$ ; two outliers  $>10$  SD excluded), indicating that extended temporal integration supports generalization. (D) No significant correlation was found between SDE-recent and learning transfer ( $r = -0.22, p = 0.12, N = 50$ ), suggesting that recent serial dependence does not predict generalization. Learning transfer was defined as the change in TDT threshold between Day 4 (final day at the first location) and Day 5 (initial day at the second location), with negative values indicating performance loss. In B-D, shaded regions denote 95% bootstrap confidence sleeves around the orthogonal regression fit.

Across the eight days of training, SDEs thus remained robust even as texture-discrimination thresholds improved markedly from Day 1 ( $126 \pm 6$  ms) to Day 8 ( $80 \pm 2$  ms;  $F(7, 343) = 41.77$ ,  $p < 0.0001$ ; Figure 4A). Biases were highly correlated across locations ( $r = 0.51$ ,  $p < 0.001$ ; Figure 4B), suggesting that the magnitude of serial dependence reflects a stable observer-specific trait consistent across retinotopic locations.

## Within-session SDE dynamics

Within-session analyses (averaged across all training days and conditions) showed that serial dependence biases remained significant throughout sessions but decreased by approximately 17% from the beginning to the end. To track such bias changes, while having a sufficient number of trials for bias analysis, each session was divided into three parts. Biases in the first third of trials (SDE-all =  $43 \pm 3\%$ ) were significantly higher than in the final third (SDE-all =  $36 \pm 4\%$ ), yielding an  $8 \pm 3\%$  reduction ( $t(49) = 2.4$ ,  $p < 0.05$ , Cohen's  $d = 0.3$ ; Figure 5A). This decrease may reflect sensory adaptation developing over the course of the session, diminishing serial dependence. The reduction was selective to distant lag history: SDE-distant decreased significantly from  $11 \pm 1\%$  to  $7 \pm 1\%$  ( $t(49) = 2.9$ ,  $p < 0.01$ , Cohen's  $d = 0.4$ ; Figure 5B, right; Supplementary Figure S2D), whereas SDE-recent remained stable ( $28 \pm 2\%$  in both segments;  $t(49) = 0.31$ ,  $p = 0.76$ ; Figure 5B, left). Notably, the 1-back bias (W1) showed a slight increase from start to end ( $2 \pm 1\%$ ;  $t(49) = 2.2$ ,  $p < 0.05$ , Cohen's  $d = 0.3$ ; Figure 5A), indicating that within-session adaptation primarily attenuates the influence of distant SDEs while leaving immediate history effects intact, or even slightly enhanced.

Observers from all conditions were combined in this analysis to increase statistical power, as separate condition-level analyses revealed the same trend (no change in SDE-recent and >30% reduction in SDE-distant), but these did not reach significance, likely due to smaller sample sizes.

## SDE differences between conditions and learning generalization

A comparison across the three experimental conditions revealed similar magnitudes of SDE-recent (dummy:  $28 \pm 3\%$ ; 1loc:  $26 \pm 2\%$ ; 2loc:  $25 \pm 2\%$ ;  $F(2,47) = 0.26$ ,  $p = 0.77$ ; Figure 6B left). In contrast, SDE-distant differed significantly between conditions (dummy:  $12 \pm 2\%$ ; 2loc:  $10 \pm 1\%$ ; 1loc:  $2 \pm 2\%$ ;  $F(2,47) = 12.47$ ,  $p < 0.001$ ; Figure 6B right). Post-hoc Tukey tests confirmed that SDE-distant was significantly lower in the 1loc condition compared to both the dummy ( $p < 0.001$ ) and 2loc ( $p < 0.01$ ) conditions, likely due to stronger sensory adaptation caused by repeated stimulation at a fixed location in the 1loc setup. In the letter discrimination (fixation control) task, which involved identical foveal stimuli across all conditions, no significant differences were observed for either SDE-recent (dummy:  $6 \pm 2\%$ ; 1loc:  $8 \pm 2\%$ ; 2loc:  $8 \pm 2\%$ ;  $F(2,47) = 0.40$ ,  $p = 0.674$ ) or SDE-distant (dummy:  $2 \pm 1\%$ ; 1loc:  $2 \pm 1\%$ ; 2loc:  $3 \pm 1\%$ ;  $F(2,47) = 0.59$ ,  $p = 0.56$ ). The smaller biases in this task likely resulted from higher overall performance levels and ceiling effects.

We next examined whether the greater learning generalization observed in the 2loc and dummy conditions (Harris et al., 2012) is linked to the stronger distant serial dependence found in those conditions (Figure 7A).

Supporting this, we found that learning transfer correlated positively with SDE-distant across observers ( $r = 0.37$ ,  $p < 0.01$ ; Figure 7B) and with SDE-distant values normalized to the 1-back effect to estimate the temporal decay constant of SDE ( $r = 0.50$ ,  $p < 0.001$ ; Figure 7C). In contrast, SDE-recent showed no positive correlation ( $r = -0.22$ ,  $p = 0.12$ ; Figure 7D), and became significantly negative when one outlier (>3 SD) was excluded ( $r = -0.32$ ,  $p < 0.05$ ,  $N = 49$ ), suggesting that recent-trial biases, being more closely tied to the current stimulus, may have a weaker and less consistent relationship with learning generalization (Figure 7A).

## Discussion

Our investigation of serial dependence in the texture discrimination task (TDT) reveals robust perceptual biases toward the orientation of previously presented targets, extending up to 10 trials back under certain conditions. Notably, these biases persisted despite randomized target orientations and significant improvements in performance across training days, suggesting that

serial dependence is a fundamental feature of visual processing, largely unaffected by task demands or learning. While our findings are based on the texture discrimination task, we expect the link between long-range serial dependence and learning generalization to extend across perceptual domains. Serial dependence and perceptual learning have been documented for numerous features including orientation, numerosity, face identity, and auditory pitch (Manassi et al., 2023 [↗](#); Lau & Maus, 2019 [↗](#); Sagi, 2011 [↗](#)), suggesting that future work could further test this link as our framework predicts. Considering the universality of learning mechanisms in the brain (Censor et al., 2012 [↗](#)), we suggest that this newly established link is not limited to visual perception but rather a general property of human behavior.

While previous studies typically report a 3-back limit for serial dependence (Fischer & Whitney, 2014 [↗](#); John-Saaltink et al., 2016 [↗](#); Lau & Maus, 2019 [↗](#); Manassi et al., 2019 [↗](#), 2023 [↗](#)), our results reveal exceptionally long memory traces, extending up to 8-back in the 2loc condition and up to 9-back in the dummy condition. Consistent with earlier work, we found that the reliability of both current and prior target stimuli affected bias magnitude. Biases increased when current targets were less visible (Ceylan et al., 2021 [↗](#); Cicchini et al., 2017 [↗](#), 2018 [↗](#); Manassi et al., 2018 [↗](#)) and when prior targets were more visible (Pascucci et al., 2019 [↗](#); Van Bergen & Jehee, 2019 [↗](#)). This pattern aligns with Bayesian models of perception (Kersten et al., 2004 [↗](#); Knill & Pouget, 2004 [↗](#)) which propose that perceptual estimates under uncertainty integrate current sensory evidence with prior information. Additionally, we observed spatial selectivity in the 2loc condition, with stronger biases when prior and current targets appeared at the same location, in agreement with previous reports (Collins, 2019 [↗](#); Fischer & Whitney, 2014 [↗](#); Fornaciai & Park, 2018 [↗](#); John-Saaltink et al., 2016 [↗](#); Manassi et al., 2019 [↗](#)). Most interestingly, in our experiments without feedback on the texture task, the experimental conditions yielding the strongest bias were also reported to enhance learning in the absence of feedback (Liu et al., 2012 [↗](#)).

Importantly, our findings reveal a functional link between serial dependence and perceptual learning. Serial dependence extends beyond transient perceptual biases and influences long-term learning, specifically, the extent to which learning generalizes to new, untrained locations. Observers trained under conditions that promote generalization ('2loc', 'dummy'; Harris et al., 2012 [↗](#)) exhibited significantly stronger and more temporally extended serial dependence from distant trial histories (4–6 back). Conversely, consistent stimulus repetition in the '1loc' condition, which promotes location-specific learning, was associated with a shorter temporal span of serial dependence, likely due to the stronger sensory adaptation (Censor & Sagi, 2009 [↗](#); Harris et al., 2012 [↗](#); Ofen et al., 2007 [↗](#)). Across individuals in all conditions, greater distant SDEs predicted greater learning transfer. These results suggest a unified mechanism in which short-term memory traces, as reflected in serial dependence, can either accumulate to support generalization or be truncated, possibly by adaptation, limiting learning to the trained context. Limited generalization is often attributed to smaller or less variable training sets in machine learning (Ying, 2019 [↗](#)), and in perceptual learning (Sagi, 2011 [↗](#)), which can lead to overfitting. A similar principle may apply here: the shorter integration window in 1loc limits the accumulation of informative variability, promoting overfitting and thus reducing generalization, thus a longer history adds little value (Tavoni et al., 2022 [↗](#)).

To our knowledge, no previous study has experimentally linked serial dependence to long-term perceptual learning. However, a theoretical framework proposed by Pascucci et al. (2019 [↗](#)) connects short-term history biases to learning mechanisms, suggesting that what appears as bias in serial dependence tasks may actually reflect the process by which the visual system updates its decision templates, the same mechanism thought to underlie perceptual learning (Doshier & Lu, 1998 [↗](#); Kuai et al., 2013 [↗](#)). Specifically, they argue that serial dependence arises from the reinstatement of previously informative sensory channels, effectively reusing feature weights that were beneficial in previous trials. Similarly, Talluri et al. (2018 [↗](#)) found that observers selectively overweight evidence consistent with prior choices, and Murai & Whitney (2021 [↗](#)), using classification-image analysis, demonstrated that serial dependence reshapes the perceptual templates applied to upcoming stimuli. Supporting this view, Urai et al. (2019 [↗](#)) fitted bounded-accumulation models and showed that choice history is best explained by a history-dependent

change in evidence accumulation (implemented as a drift bias), rather than merely shifting the starting point of the decision process (criterion bias). This result is in agreement with our RT analysis showing that distant SDEs are RT-independent, a marker of drift bias (Dekel & Sagi, 2020). In addition, the recent trials seem to introduce criterion shifts (starting point bias in the drift diffusion model), indicated by the larger biases found for fast RTs. Together, these findings suggest that serial dependence directly alters how sensory information is weighted and interpreted. We suggest that these updated decision templates subservise perceptual learning.

The persistence of serial dependence during eight days of training with random stimulus sequences, where it does not contribute to online performance, but rather increases decision noise, suggests that the assumption of environment stability is hardwired into the brain, or that these biases may serve a broader function beyond optimizing immediate performance. Consistent with our approach, recent models reframe serial dependence as a memory-driven phenomenon, not as an optimal inference about the external world (Barbosa & Compte, 2020), but as a consequence of internal mechanisms shaped by how recent perceptual states are encoded and maintained over time (Kalm & Norris, 2018).

Our findings reveal a functional/mechanistic dissociation between short- and long-range serial dependence. Only recent SDEs were modulated by reaction time, presenting stronger biases with faster responses, suggesting that these biases are due to shifts in decision criteria (Dekel & Sagi, 2020). In contrast, distant SDEs were found to be RT independent, suggesting that these biases are a result of neuronal reweighting (Dekel & Sagi, 2020). Importantly, only distant SDEs predicted learning transfer, while recent SDEs remained stable across conditions and were unrelated to generalization. The within session dynamics showed distant SDEs, but not recent SDE, to decline with training, thus effectively reducing the SDE range, consistent with the increased learning specificity observed in perceptual learning with extensive learning (Sagi, 2011). This pattern suggests a functional distinction: recent biases may relate more to prior stimulus statistics, whereas temporally extended biases may support the integration of sensory evidence required for efficient perceptual learning. Previous studies also point to distinct timescales in serial dependence. For example, Lieder et al. (2019) showed that perceptual biases reflect processes operating over different timescales that vary across clinical populations: individuals with ASD rely less on recent trials but show intact long-term integration, whereas individuals with dyslexia exhibit the opposite pattern. Thus, in the absence of adaptation, we expect learning in ASD to generalize, as indeed was recently found (Harris et al., 2015). Fritsche et al. (2020) proposed a model in which perceptual history influences current biases through both short-term Bayesian decoding and longer-term efficient encoding, aligning with our observed dissociation between recent and distant SDEs. While these converging findings support distinct mechanisms for recent and distant SDEs, our correlational approach cannot definitively establish causality, and targeted experimental manipulations would further strengthen these interpretations.

Our findings offer a new insight into the mechanisms of perceptual learning. While traditional theories explain learning specificity through local changes at the site of target encoding (Karni & Sagi, 1991), the formation of location-specific decision templates (Doshier & Lu, 1999), or both (Karni & Sagi, 1993; Watanabe & Sasaki, 2015), we propose a unified mechanism. Specifically, we suggest an account based on a single decision template that learns the discrimination task by classifying neuronal response features as signaling vertical or horizontal texture targets. These templates generalize across retinal locations of equal eccentricity but not across locations with different eccentricities (Harris & Sagi, 2018). However, when trained with targets in a fixed location, the decision template may become overfitted to features that are specific to that location, limiting generalization (Sagi, 2011). For learning to generalize, multiple samples (trials) must be integrated over time to filter out local noise. Our results show that decision biases are integrated linearly over trials, suggesting efficient temporal integration over many trials in conditions that support learning generalization. In contrast, reduced integration, due to adaptation, or increased inhibition, may produce classifiers that rely on location-specific noise (Mollon & Danilova, 1996). We suggest that previous reports of learning generalization can be explained by a modulation of temporal integration. This includes short training phases that are stopped before adaptation takes

over, showing generalization to other retinal locations (Censor & Sagi, 2009; Karni & Sagi, 1993), short pre-training phases enabling generalization across visual tasks (Zhang et al., 2010) and other paradigms that effectively reduce sensory adaptation (reviewed in Sagi, 2011) and by that allow serial dependence to accumulate. Our findings thus provide empirical support for a unified mechanism that governs both specific and generalized learning through modulation of temporal integration. We further speculate that the integration window is affected by the balance between excitation and inhibition (E/I balance) in the visual cortex, shown to affect learning stabilization in TDT (Shibata et al., 2017; Tamaki et al., 2020). Computational models implementing trial-by-trial reweighting (e.g. Petrov et al., 2005) with adaptation dependent reweighting, can potentially account for SDE decay profiles and their relationship to generalization, providing quantitative predictions for future experiments (or for already existing experimental data reanalyzed for SDE). Weights update dynamics may affect network flexibility and generalization. To account for the present results within this general framework, we assume that consistent stimulus repetition (triggering sensory adaptation) stabilizes learning by lowering the gain of network update. Reduced sensory adaptation (disinhibition) allows for increased plasticity producing network dynamics allowing for faster adjustment to new stimuli. This update mechanism is expected to introduce serial dependencies with a temporal scale defined by E/I balance. To test the plausibility of this approach, we constructed a simple computational model of learning presented in Supplementary results. This toy model, based on modelling learning in volatile environments (Piray and Daw, 2020), predicts serial dependence produced by template update, with a magnitude correlated with generalization. An important property of this model is the continuous plasticity, that is the system does not stop updating its templates. Attractive serial dependence emerges as an immediate consequence of ongoing template learning: informative (high SOA, easy) trials selectively update orientation templates, and subsequent ambiguous (low SOA, noisy) trials read out these updated templates, producing an attractive bias without requiring explicit feedback or top-down control. In addition, we may consider biases due to updated priors, as commonly assumed in the SDE literature. This positions serial dependence as a signature of the same flexibility mechanisms that support adaptive learning in nonstationary environments.

In summary, we show that long-range serial dependence predicts learning transfer, supporting the view that short-term memory contributes directly to long-term learning. By connecting serial dependence with learning, our findings bridge a key theoretical gap and suggest that the integration of past experience plays a crucial role in determining the specificity or generalization of learning.

## Methods

### Overview

We simulated a nonstationary 1D categorization task with online template learning and quantified lagged sequential dependence effects (SDE) on current **hard A/B** trials near the true category boundary. Simulations were Monte Carlo over independent random seeds and additionally included a third **noise (dummy)** trial type embedded in the stream.

### Environment and trial generation

Each simulation comprised  $T = 100,000$  trials with regime switches governed by a Bernoulli hazard  $H = 1/10000$  per trial. Regime 1 had true category means  $(\mu_A, \mu_B) = (-1, +1)$ ; regime 2 had  $(\mu_A, \mu_B) = (0, +2)$ . On each trial, with probability  $\pi_N = 0.25$  the sample was a **noise** trial generated as  $x_t \sim \mathcal{N}(\mu_N, \sigma_N^2)$  with  $\mu_N = 0$  and  $\sigma_N = 0.1$ ; otherwise the trial was an A/B trial with category B drawn with base rate  $\pi_B = 0.5$  and sample generated as  $x_t \sim \mathcal{N}(\mu_A^*(t), \sigma_t^2)$  or  $x_t \sim \mathcal{N}(\mu_B^*(t), \sigma_t^2)$ . A reliability schedule controlled  $\sigma_t$  for A/B trials: a warm-up period of  $t_{\text{Warm}} = 100$  trials used “easy” trials only  $\sigma_{\text{Easy}} = 0.5$ ; thereafter each A/B trial was easy with probability  $p_{\text{Easy}} = 0.25$  and hard otherwise, using  $\sigma_{\text{Hard}} = 1.0$ . Easy/hard tags were defined for A/B trials only (noise trials excluded). Noise trials were part of the temporal stream and counted toward lags; in the default configuration they updated the learner. Noise trials simulated dummy trials and were not present when the standard 1loc condition was simulated.

## Distance-to-boundary binning

For each trial we computed the true midpoint boundary  $m^*(t) = \frac{1}{2}(\mu_A^*(t) + \mu_B^*(t))$  and distance  $d_t = |x_t - m^*(t)|$ . Trials were binned into **near/mid/far** using thresholds  $\text{thr}_1 = 0.5$  and  $\text{thr}_2 = 1.0$  (near if  $d_t \leq \text{thr}_1$ , mid if  $\text{thr}_1 < d_t \leq \text{thr}_2$ , far if  $d_t > \text{thr}_2$ ). Primary SDE analyses focused on *current* hard A/B trials in the near bin.

## Learner (online template model)

The observer maintained estimates of A and B template means  $\hat{\mu}_A(t), \hat{\mu}_B(t)$ , scalar uncertainties  $P_A(t), P_B(t)$ , and (optionally) a dynamic prior  $p_B(t)$ . On each trial it computed a soft responsibility  $r_t = P(B | x_t)$  from Gaussian-shaped likelihoods centered on the current templates with fixed sensory scale  $\sigma_{\text{sense}} = 0.6$  and prior  $p_B$ , and generated a binary response B when  $(r_t > 0.5)$ , or A otherwise. Template updates were Kalman-like: uncertainties were inflated by a process-noise term  $Q_t$ , responsibility-weighted effective observation noises were formed, Kalman gains computed, and both templates were updated proportional to prediction errors. Update magnitude was modulated by a confidence gate  $c_t$  derived from  $|r_t - 0.5|$  (parameters  $\text{confMin} = 0.05, \text{confPower} = 2$ ). Flexibility was controlled by a VKF-like volatility estimate: a latent volatility state  $v_t$  tracked an exponential moving average of squared prediction errors with forgetting  $l = 0.05$ , and process noise was set to  $Q_t = \text{clip}(Q_0 + g_v v_t, [0, Q_{\text{max}}])$  with  $Q_{\text{max}} = 0.2$ . Baseline observation noise was  $R_{\text{base}} = 0.1$  (scaled by  $\sigma_{\text{sense}}^2$ ). Learning occurred on all trials, including noise. A sticky-prior module (disabled for template-shift-only analyses) updated  $p_B$  toward  $r_t$  with  $\text{etaPrior} = 0.5$  and relaxed it toward 0.5 with time constant  $\text{tauPrior} = 3$ . Initial learner state was  $\hat{\mu}_A(0) = -0.2, \hat{\mu}_B(0) = +0.2, P_A(0) = P_B(0) = 1.0, p_B(0) = 0.5$ , and  $v_0 = 0$ .

## Monte Carlo and analysis windows

We simulated  $n_{\text{Seeds}} = 100$  independent environments and ran the learner on each. All analyses used a burn-in period  $t_{\text{Burn}} = t_{\text{Warm}} + 1000 = 1100$  trials to exclude initialization transients. Lag analyses used maximum lag  $L_{\text{max}} = 10$  and required at least  $\text{minCount} = 200$  samples per condition.

## Sequential dependence metric

For each lag  $\ell$ , we computed the bias in B reports on a set of *current* trials conditioned on the true label of the trial  $\ell$  steps back, restricted to cases where that previous trial was **easy A/B**. Specifically, we estimated  $b_B(\ell) = 2P(\text{report B} | \text{current mask, prev easy true B}) - 1$  and  $b_A(\ell) = 2P(\text{report B} | \text{current mask, prev easy true A}) - 1$ , and combined them as  $\Delta P_B(\ell) = \frac{1}{2}(b_B(\ell) - b_A(\ell))$ . This analysis was carried out on current **hard A/B** trials in the **near** bin, with SEM computed across seeds (decisions on noisy, dummy, trials were excluded).

## Mechanistic boundary-step coupling

To link SDE to learning-driven boundary motion, we defined the learned midpoint  $m_t = \frac{1}{2}(\hat{\mu}_A(t) + \hat{\mu}_B(t))$  and boundary step  $\Delta m_t = m_t - m_{t-1}$ . We measured  $\Delta m_t$  on easy A/B trials split by true label, and quantified coupling by binning  $\Delta m_{t-1}$  from previous easy trials and computing  $P(\text{report B at } t)$  on the immediately subsequent hard A/B trial.

## Adaptation time after regime switches

Flexibility was quantified as relearning time following regime switches using the template error  $E_t = |\hat{\mu}_A(t) - \mu_A^*(t)| + |\hat{\mu}_B(t) - \mu_B^*(t)|$ . Switches were separated by at least 5000 trials. For each switch, baseline error was the mean error in the pre-window; relearning time was the first post-switch time when error returned to within  $\text{tolFrac} = 0.15$  of baseline, averaged across switches and seeds.

## Parameter sweep

To characterize the bias–adaptation tradeoff, we ran a grid sweep over volatility parameters  $Q_0 \in \{0.5 \cdot 10^{-5}, 10^{-4}, 2 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}\}$  and  $g_V \in \{0.5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 4 \cdot 10^{-3}\}$ , recomputing mean adaptation time and SDE at lags  $\{1, 4, 10\}$  (default near-bin hard A/B) for each parameter pair; points were visualized as adaptation time versus SDE magnitude, color-coded by mean effective process noise  $\langle Q_t \rangle$ .

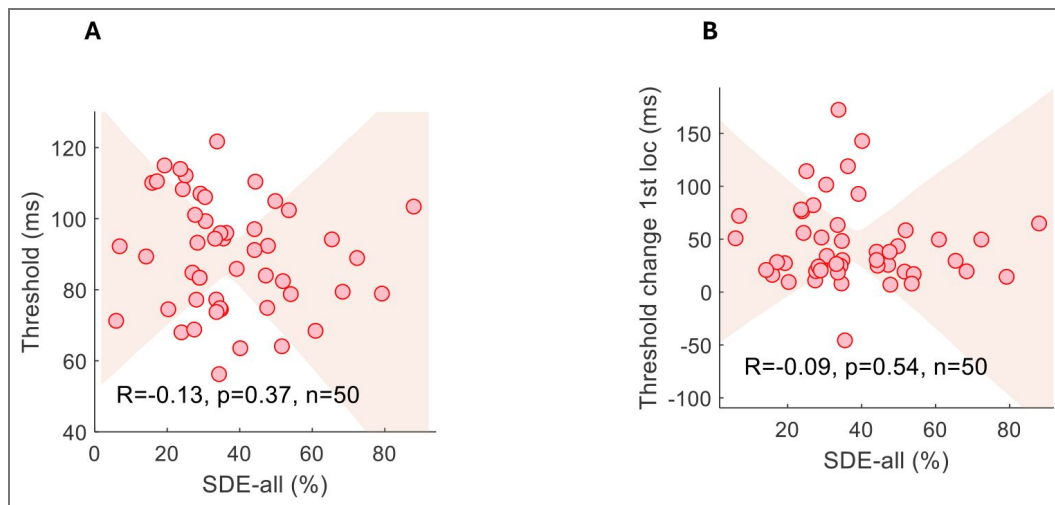
## Compact parameter list (defaults)

Monte Carlo: seedBase 42,  $N = 100$ . Environment:  $T = 100,000$ ,  $t_{\text{warm}} = 100$ , hazard  $H = 1/10000$ ,  $\pi_B = 0.5$ , regimes  $(\mu_A, \mu_B) = (-1, +1)$  and  $(0, +2)$ . Reliability:  $p_{\text{easy}} = 0.25$ ,  $\sigma_{\text{easy}} = 0.5$ ,  $\sigma_{\text{hard}} = 1.0$ . Noise trials:  $\pi_N = 0.25$ ,  $\mu_N = 0$ ,  $\sigma_N = 0.1$ . Analysis:  $t_{\text{burn}} = 1100$ ,  $L_{\text{max}} = 10$ , minCount=200, bins  $\text{thr}_1 = 0.5$ ,  $\text{thr}_2 = 1.0$ . Learner: sigSense=0.6, R\_base=0.1, confidence on (confMin=0.05, confPower=2), VKF on, lambda=0.05,  $Q_0=10^{-5}$ ,  $Q_{\text{max}}=0.2$ , initial  $\hat{\mu}_A = -0.2$ ,  $\hat{\mu}_B = +0.2$ ,  $P_0 = 1$ ,  $v_0 = 0$ ,  $p_{B,0} = 0.5$ . Sticky prior for optional runs: etaPrior=0.5, tauPrior=3 (disabled for template-shift-only analyses).  $g_V = 0.005$  for dummy,  $g_V = 0$  for 1loc.

## Supplementary material: Tables and Figures

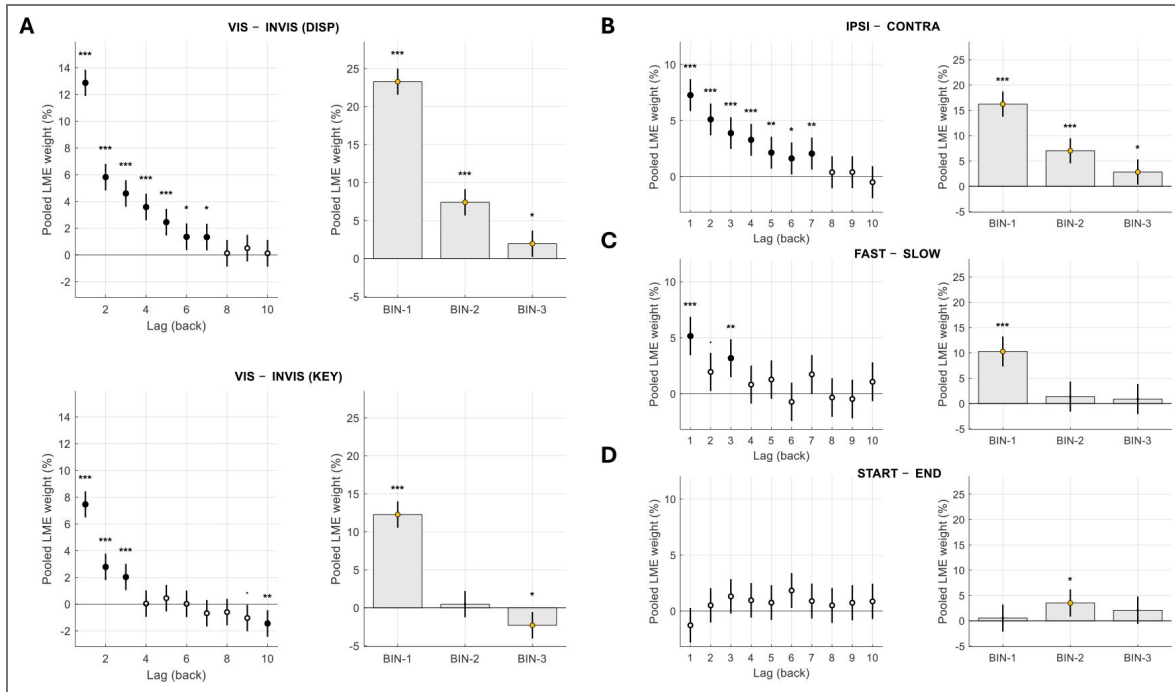
| Condition | Group            | N | Comments  |
|-----------|------------------|---|---|
| 1loc      | 1loc-standard    | 7 |   |
|           | 1loc-dark        | 7 | Double inter-trial period                         |
| 2loc      | 2loc-standard    | 7 |   |
|           | 2loc-long        | 7 | Twice the number of trials                        |
| Dummy     | dummy-standard   | 7 | Background of horizontal bars                     |
|           | dummy-90reBck    | 5 | Background of vertical bars                       |
|           | dummy-90reTarget | 5 | Background of diagonal bars (angled at 45°)       |
|           | dummy-local      | 5 | Local background with only 5 × 5 texture elements |

**Table S1.** Group assignments in each condition



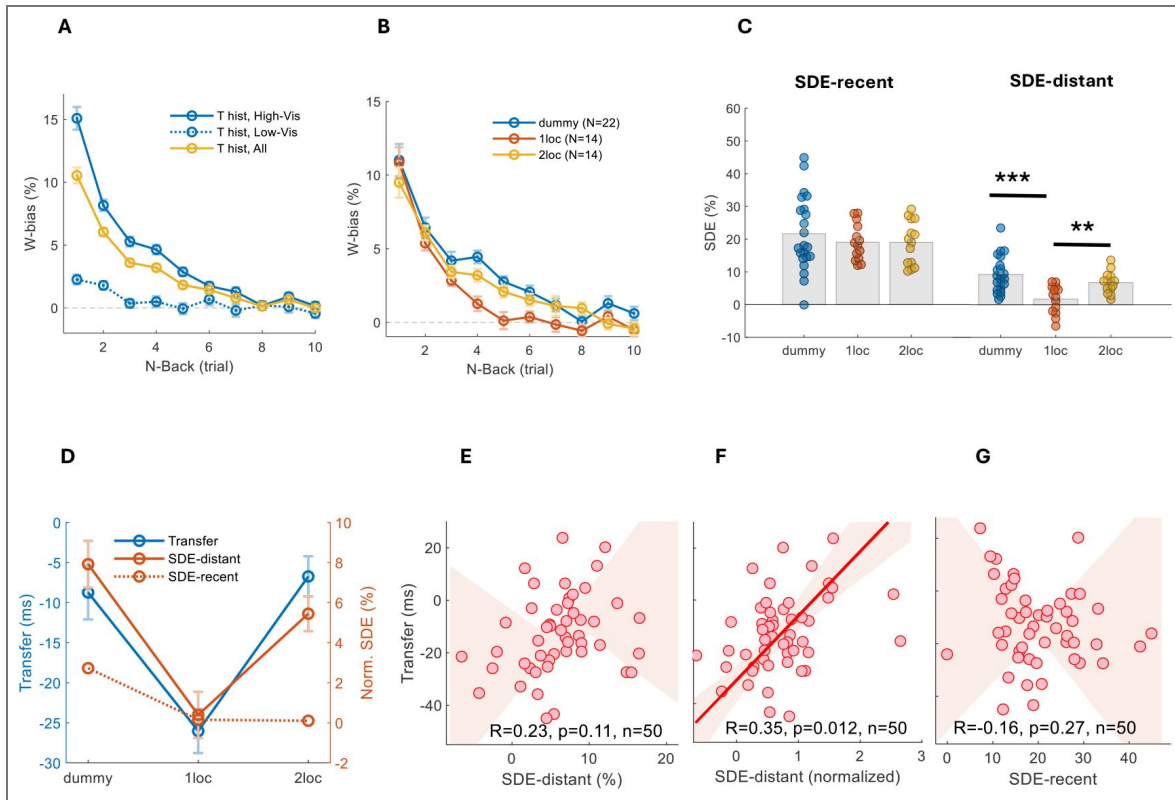
**Figure S1. Relationship between TDT Threshold and SDEs**

(A) Correlation between average TDT thresholds (Days 1–8) and SDE-all across observers ( $n = 50$ ). No significant relationship was found ( $r = -0.13$ ,  $p = 0.37$ ). (B) Correlation between threshold improvements at the first location (Day 1 to day 4) and SDE-all. No significant relationship was found ( $r = -0.09$ ,  $p = 0.54$ ), suggesting that the magnitude of serial dependence does not predict the overall amount of perceptual learning.



**Figure S2. Serial-dependence contrasts from pooled LME estimates**

Each subfigure shows SDE across individual lags (left) and lag bins (right). For each contrast, lag-specific history weights from the linear mixed-effects (LME) models were computed as the difference between conditions over the ten preceding trials (lags 1–10). Differences were computed either within the same model (VIS – INVIS, IPSI – CONTRA) or between independent model fits trained on matched subsets of trials (FAST – SLOW, START – END). Estimates were pooled across observers using inverse-variance weighting (grand-pooled across groups for VIS – INVIS; cross-fit and pooled across observers for FAST – SLOW and START – END; IPSI – CONTRA computed for loc2). Values are pooled contrast estimates ( $\Delta \pm SE$ , % bias). Significance is based on FDR-corrected  $q$ -values (Benjamini–Hochberg across lags). Bin panels summarize BIN1 (1–3 back), BIN2 (4–6 back), BIN3 (7–9 back). **(A) VIS – INVIS.** SDEs were reduced after low-visibility trials, with a larger reduction for display-based (target) than for key-based (response) histories; key-driven SDEs showed a small significant increase at distant lags for invisible trials. **(B) IPSI – CONTRA.** SDEs were stronger for ipsilateral than contralateral locations, indicating spatial specificity. **(C) FAST – SLOW.** Trials were split into fastest and slowest reaction-time quartiles per observer and day. Biases were stronger for fast trials, particularly at recent lags (SDE-recent). **(D) START – END.** SDEs computed for the first vs. last third of each session showed an overall decline toward session end, with slight increase in the immediate (1-back) bias. The reduction reached significance only for BIN2 (SDE-distant). Error bars denote  $\pm 1$  SEM across observers. Asterisks indicate FDR-corrected significance ( $***q < .001$ ,  $**q < .01$ ,  $*q < .05$ ).



**Figure S3. Serial dependence effects using all prior-trial history (unfiltered).**

(A) W-bias as a function of n-back lag, comparing high-visibility (solid blue), low-visibility (dashed blue), and all prior trials (yellow). Data pooled across conditions. (B) W-bias decay across n-back lags for each experimental condition (dummy, 1loc, 2loc) using all prior trials. As with the high-visibility prior analysis, biases decayed more rapidly in the 1loc condition, while persisting further back in trial history for the 2loc and dummy conditions. (C) Individual SDE values by condition for SDE-recent (left) and SDE-distant (right). Bars indicate group means; dots represent individual observers. As with the high-visibility prior analysis, recent SDEs were consistent across conditions (dummy:  $29 \pm 3\%$ ; 1loc:  $29 \pm 2\%$ ; 2loc:  $28 \pm 2\%$ ;  $F(2,47) = 0.08$ ,  $p = 0.92$ ), whereas distant SDEs were significantly stronger in the 2loc and dummy conditions compared to the 1loc condition (dummy:  $12 \pm 2\%$ ; 1loc:  $3 \pm 1\%$ ; 2loc:  $11 \pm 1\%$ ;  $F(2,47) = 11.27$ ,  $p = 0.001$ ). Post-hoc Tukey tests confirmed that SDE-distant was significantly lower in the 1loc condition compared to both the dummy ( $***p < 0.001$ ) and 2loc ( $**p < 0.01$ ) conditions. (D) Group-level comparison of learning transfer (blue), SDE-distant (solid red), and SDE-recent (dotted red) across the three experimental conditions (dummy, 1loc, 2loc). Transfer and SDE values are plotted on separate axes, with SDE measures normalized by subtracting the mean of the 1loc condition. Conditions showing greater learning generalization (dummy, 2loc) also exhibited stronger SDE-distant effects. In contrast, SDE-recent was relatively constant across conditions, suggesting that generalization was primarily linked to distant serial dependence. (E) Across observers, learning transfer was not significantly correlated with SDE-distant ( $r=0.23$ ,  $p=0.11$ ,  $N=50$ ), suggesting that using all prior-trial history reduces the effect. (F) SDE-distant values were normalized to the 1-back effect to estimate the temporal decay constant of SDE, reflecting how long biases persisted across trials. Observers with longer decay constants showed greater learning transfer ( $r=0.35$ ,  $p=0.012$ ,  $N=50$ ), similar to the high-visibility prior analysis. (G) No significant correlation was found between SDE-recent and learning transfer ( $r=-0.16$ ,  $p=0.27$ ,  $N=50$ ), suggesting that recent serial dependence does not predict generalization. In E-G, shaded regions denote 95% bootstrap confidence sleeves around the orthogonal regression fit.

## Supplementary material: modelling

Here we describe a simple model of learning that generates persistent serial dependence (SDE) as a consequence of weight update. There is no intention of fitting the experimental data but rather showing that SDE can arise as a consequence of weights update in a learning model. We here describe the learning mechanism, followed by model equations and simulations.

In the experiments simulated here, on each trial observers are visually presented with one of two stimulus categories, Vertical or Horizontal. Targets are masked by noise of variable impact, determined by Stimulus Onset Asynchrony (SOA) between the target and mask stimuli. On each trial, observers decide whether the presented stimulus is Vertical or Horizontal. There was no response feedback in the experiments and in the simulations. The main results to account for are the long range SDE and the correlation between SDE and learning transfer.

In the simulations, observers are assumed to maintain two online category templates,  $\mu_A$  and  $\mu_B$ , updated from trial-by-trial sensory samples via responsibility-weighted, Kalman-like template updates. Updates follow a variant of the Volatile Kalman Filter (VKF; Piray & Daw, 2020). In our simulations the categories are orientations: the observer keeps two running estimates of what Vertical (category A) and Horizontal (category B) targets “look like” along a 1-D decision axis (orientation evidence in our experiments). These internal templates correspond to current estimates of the category means,  $\mu_A$  and  $\mu_B$ .

On each trial, the visual system generates an internal sensory response  $x_t$ , added with stimulus-independent sensory noise  $\sigma$ . Using the two templates and the current category prior, a posterior responsibility can be computed

$$r_t = P(B | x_t)$$

i.e., how strongly the current sample supports category B rather than A. This responsibility is used for:

- Decision: report B if  $r_t > 0.5$ .
- Learning: update both templates, but with responsibility-dependent weights. If  $r_t \approx 1$ , the B-template is updated strongly and the A-template minimally; if  $r_t \approx 0.5$ , both updates are small; if  $r_t \approx 0$ , the A-template is updated strongly and the B-template minimally.

There are easy and hard trials defined by SOA: easy trials (high SOA, the noisy mask does not interfere with the target stimulus) tend to generate samples  $x_t$  farther from the decision boundary, yielding more extreme responsibilities ( $r_t$  near 0 or 1), whereas hard trials (low SOA, target corrupted by noise) yield samples closer to the decision boundary with  $r_t$  near 0.5. As a result, the internal state (templates, and therefore the implied boundary between them) shows jumps followed by persistence: informative/easy samples produce larger template shifts, while ambiguous/hard samples produce little change.

Serial dependence follows directly. An easy trial shifts the internal boundary; subsequent hard trials are ambiguous and therefore sensitive to the exact positioning of the boundary. Because hard trials do not strongly update the templates, the boundary persists and the bias extends across multiple lags. In the simulations this link is verified by computing the boundary shift

$$\Delta m_t = m_t^{\text{post}} - m_t^{\text{pre}} :$$

and showing that  $\Delta m_t$  differs after EASY-A versus EASY-B trials. Moreover, for EASY → HARD pairs,  $\Delta m_{t-1}$  predicts the probability of reporting B on the next hard trial.

Of interest here is that the same mechanism that generates SDE also determines flexibility (generalization). In VKF-like updates, volatility enters as a process-noise term  $Q_t$  that controls the Kalman gains (see below ‘Volatility’). When mismatch accumulates (operationalized below as a large prediction error  $(x_t - \hat{x}_t)^2$ ; e.g., after a rotation/location switch),  $Q_t$  increases, gains increase, and the templates move faster toward the new mapping—yielding rapid relearning, considered here as learning generalization if very fast (e.g. ‘dummy’ condition). When volatility is depressed (e.g. ‘1loc’ condition), gains are smaller and the template state is more conservative: boundary

shifts produced by informative events are smaller (weaker SDE, bidirectional), and re-learning after a change is slower. *The positive SDE observed with high volatility is a result of the high update gain that tends to overshoot the updated template behind its expected ‘true’ value. Thus, SDE and re-learning are two observable consequences of the same underlying state-update machinery—one read out as a bias on ambiguous trials, the other read out as the speed of convergence after a mapping change.*

In addition, we have a sticky prior over categories. Instead of keeping the prior  $p_B$  fixed at 0.5, the observer maintains an online estimate of how likely category B is: after each trial, the prior shifts slightly toward the current posterior  $r_t$  and then relaxes back toward 0.5. This, as described in the SDE literature, provides a source of serial dependence that biases inference directly (via prior odds), even in the absence of template shifts.

## Technical description

### State variables

- Template means:  $\mu_{A,t}$ ,  $\mu_{B,t}$
- Template uncertainties:  $P_{A,t}$ ,  $P_{B,t}$
- Category prior:  $p_{B,t}$
- Volatility state:  $v_t$

### Trial order

At trial  $t$ : (1) observe  $x_t$ ; (2) compute  $r_t$  using the pre-update templates and prior; (3) decide A or B; (4) update volatility; (5) update templates; (6) update sticky prior (if enabled).

### Decision (uses pre-update templates)

$$\begin{aligned} \mu_{k,t}^{\text{pre}} &= \mu_{k,t-1}^{\text{post}} \\ m_t^{\text{pre}} &= \frac{\mu_{A,t}^{\text{pre}} + \mu_{B,t}^{\text{pre}}}{2}. \end{aligned}$$

### Inference

Likelihoods use fixed internal  $\sigma$ :

$$L_A(t) \propto \exp \left[ -\frac{1}{2} \left( \frac{x_t - \mu_{A,t}^{\text{pre}}}{\sigma} \right)^2 \right], \quad L_B(t) \propto \exp \left[ -\frac{1}{2} \left( \frac{x_t - \mu_{B,t}^{\text{pre}}}{\sigma} \right)^2 \right]$$

Posterior responsibility:

$$r_t = \frac{p_{B,t} L_B(t)}{(1-p_{B,t}) L_A(t) + p_{B,t} L_B(t)}$$

Decision: report B if  $r_t > 0.5$ .

### Confidence gate

To reduce learning from ambiguous stimuli, we optionally apply a confidence gate based on posterior distance from indifference:

$$r_t - 0.5 \mid$$

A bounded confidence gate is:

$$c_t = \min \left\{ 1, \max \left( c_{\min} (2 |r_t - 0.5|)^k \right) \right\}$$

This downweights learning from ambiguous (dummy) trials ( $r_t \approx 0.5$ ).

### Volatility (flexibility)

Prediction:

$$\hat{x}_t = (1 - r_t) \mu_{A,t}^{\text{pre}} + r_t \mu_{B,t}^{\text{pre}}$$

Volatility update ( $v_t$  increases when new observation is far from the predicted,  $\hat{x}_t$ ):

$$v_t = (1 - \lambda)v_{t-1} + \lambda(x_t - \hat{x}_t)^2$$

Process noise:

$$Q_t = \text{clip}(Q_0 + g_V v_t, 0, Q_{\max})$$

## Template updating

Prediction uncertainty (updated by process noise):

$$P_{k,t}^{\text{pre}} = P_{k,t-1}^{\text{post}} + Q_t$$

Observation uncertainty (fixed):

$$R = R_{\text{base}}\sigma^2$$

Responsibility-weighted effective noise:

$$R_{A,t} = \frac{R}{1-r_t}, \quad R_{B,t} = \frac{R}{r_t}$$

Kalman gains:

$$K_{A,t} = \frac{P_{A,t}^{\text{pre}}}{P_{A,t}^{\text{pre}} + R_{A,t}}, \quad K_{B,t} = \frac{P_{B,t}^{\text{pre}}}{P_{B,t}^{\text{pre}} + R_{B,t}}$$

Updates:

$$\begin{aligned} \mu_{A,t}^{\text{post}} &= \mu_{A,t}^{\text{pre}} + c_t K_{A,t} (x_t - \mu_{A,t}^{\text{pre}}) \\ \mu_{B,t}^{\text{post}} &= \mu_{B,t}^{\text{pre}} + c_t K_{B,t} (x_t - \mu_{B,t}^{\text{pre}}) \end{aligned}$$

Boundary step:

$$\Delta m_t = m_t^{\text{post}} - m_t^{\text{pre}}$$

Large mismatch  $\rightarrow$  larger  $v_t \rightarrow$  larger  $Q_t \rightarrow$  faster re-learning.

## Sticky prior dynamics

The prior is updated as a leaky average of recent posteriors:

History integration:

$$p_{B,t}^{\text{raw}} = (1 - \eta)p_{B,t-1} + \eta r_t$$

Relaxation toward neutrality:

$$p_{B,t} = w p_{B,t}^{\text{raw}} + (1 - w) \cdot 0.5$$

Where

$$w = e^{-1/\tau}$$

So:

$\eta$  controls how strongly the current posterior influences the next prior.

$\tau$  controls how slowly the prior relaxes back toward 0.5.

This prior is used in the next trial's responsibility calculation.

## Serial dependence mechanisms

Serial dependence in this model can arise from two coupled mechanisms:

- A. Template-shift mechanism. Easy trials induce nonzero boundary steps  $\Delta m_t$ , while hard trials do not strongly overwrite the boundary. Because  $m_t^{\text{pre}}$  is autocorrelated, conditioning on a past easy A vs B trial predicts current boundary position, biasing ambiguous decisions. Volatility  $Q_t$  contributes to serial dependence because it sets the Kalman gains: higher  $Q_t$  increases the amplitude of easy-trial boundary steps (stronger bias).
- B. Sticky-prior mechanism. The prior  $p_{B,t}$  integrates recent posteriors and directly biases the next trial's inference via prior odds:

$$\text{odds}_t = \frac{p_{B,t}}{1-p_{B,t}}$$

Even if templates were frozen, this would generate serial dependence. When both mechanisms are active, SDE reflects both persistent boundary shifts and prior drift.

## Simulation results

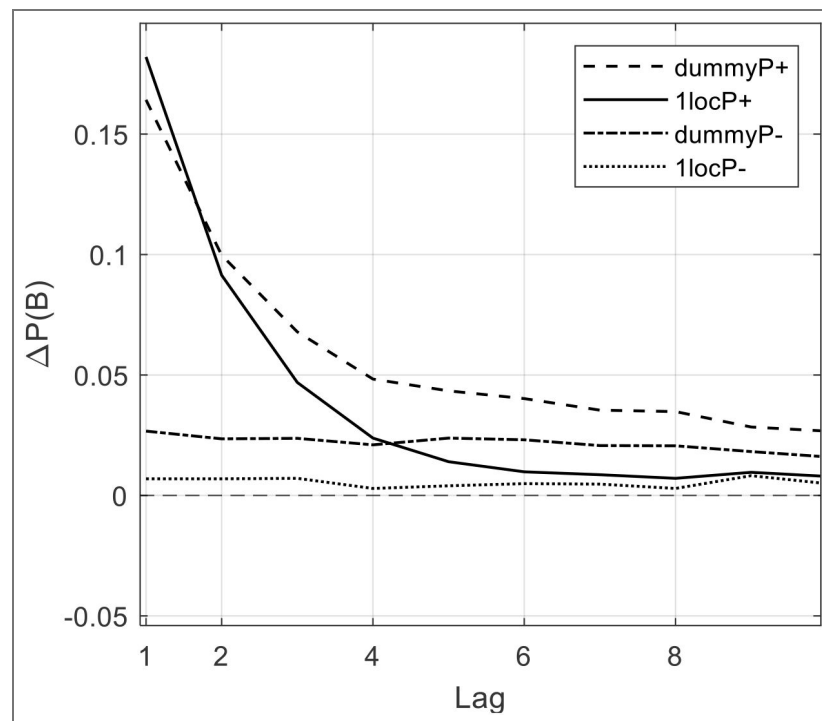
There was no attempt to fit the data, but rather to demonstrate the similarity between model behavior and that observed in the experiments. The main interest here is in showing the ability of the model to imitate the experimental results from the ‘lloc’ and the ‘dummy’ conditions, when controlling volatility ( $Q_0, g_V$ ). Simulations followed the scheme presented above. There were 100 experiments of 100K trials with easy, hard and noisy/dummy (mean=0) trials.

## Data availability

The data that support the findings of this study will be available from the corresponding author upon request.

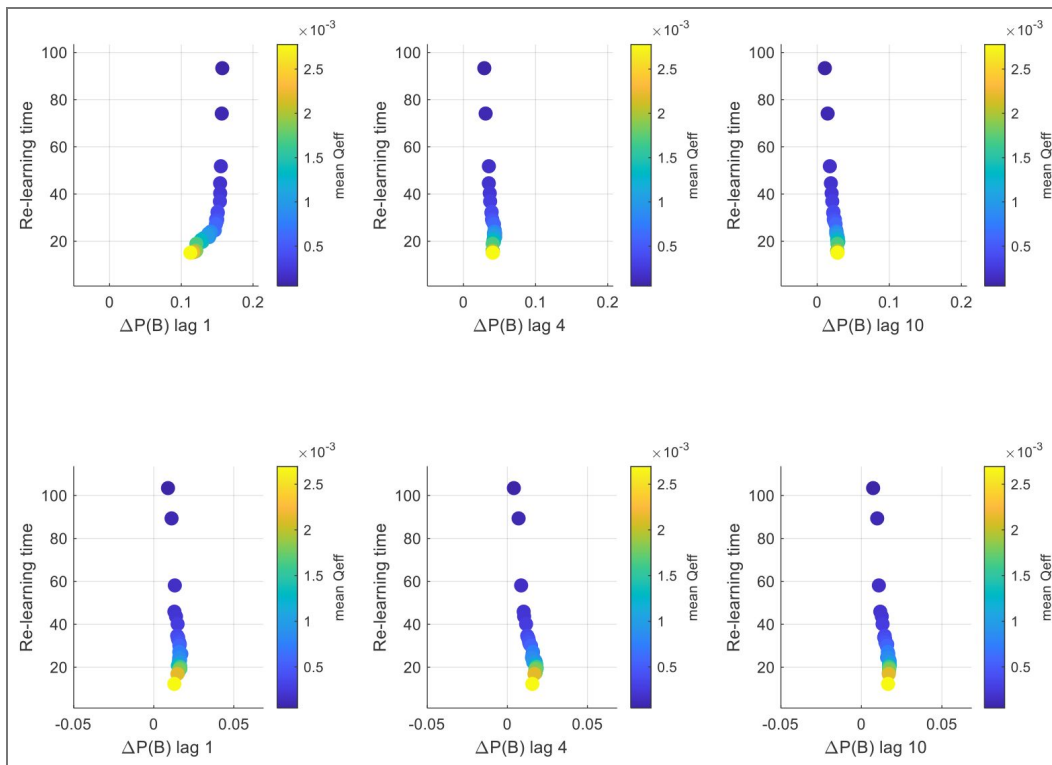
## References

- Abrahamyan A.**, Silva L. L., Dakin S. C., Carandini M., Gardner J. L. (2016) Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences* **113**:E3548-E3557 <https://doi.org/10.1073/pnas.1518786113> | PubMed
- Barbosa J.**, Compte A. (2020) Build-up of serial dependence in color working memory. *Scientific Reports* **10**:10959 <https://doi.org/10.1038/s41598-020-67861-2> | PubMed
- Braun A.**, Urai A. E., Donner T. H. (2018) Adaptive history biases result from confidence-weighted accumulation of past choices. *Journal of Neuroscience* **38**:2418-2429 <https://doi.org/10.1523/jneurosci.2189-17.2017> | PubMed
- Censor N.**, Sagi D. (2009) Global resistance to local perceptual adaptation in texture discrimination. *Vision Research* **49**:2550-2556 <https://doi.org/10.1016/j.visres.2009.03.018> | PubMed
- Censor N.**, Sagi D., Cohen L. G. (2012) Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience* **13**:658-664 <https://doi.org/10.1038/nrn3315> | PubMed
- Ceylan G.**, Herzog M. H., Pascucci D. (2021) Serial dependence does not originate from low-level visual processing. *Cognition* **212**:104709 <https://doi.org/10.1016/j.cognition.2021.104709> | PubMed
- Cheng Y.-A.**, Sanayei M., Chen X., Jia K., Li S., Fang F., Watanabe T., Thiele A., Zhang R.-Y. (2025) A neural geometry approach comprehensively explains apparently conflicting models of visual perceptual learning. *Nature Human Behaviour* **9**:1023-1040 <https://doi.org/10.1038/s41562-025-02149-x> | PubMed
- Cicchini G. M.**, Mikellidou K., Burr D. (2017) Serial dependencies act directly on perception. *Journal of Vision* **17**:6 <https://doi.org/10.1167/17.14.6> | PubMed
- Cicchini G. M.**, Mikellidou K., Burr D. C. (2018) The functional role of serial dependence. *Proceedings of the Royal Society B* **285**:20181722 <https://doi.org/10.1098/rspb.2018.1722> | PubMed
- Collins T.** (2019) The perceptual continuity field is retinotopic. *Scientific Reports* **9**:18841 <https://doi.org/10.1038/s41598-019-55134-6> | PubMed
- Dekel R.**, Sagi D. (2020) Perceptual bias is reduced with longer reaction times during visual discrimination. *Communications Biology* **3**:1-12 <https://doi.org/10.1038/s42003-020-0786-7> | PubMed
- Doshier B. A.**, Lu Z.-L. (1999) Mechanisms of perceptual learning. *Vision Research* **39**:3197-3221 [https://doi.org/10.1016/s0042-6989\(99\)00059-0](https://doi.org/10.1016/s0042-6989(99)00059-0) | PubMed
- Doshier 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**:13988-13993 <https://doi.org/10.1073/pnas.95.23.13988> | PubMed



**Figure S4.** Simulated bias as a function of lag (trials) for four conditions: (1) high volatility with sticky prior, simulating the dummy condition, (2) low volatility with sticky prior, simulating the 1loc condition, and (3) high volatility without sticky prior, and (4) low volatility without sticky prior.

Condition 3 (dummyP-) reflects pure template-learning SDE (i.e., serial dependence arising from template updates alone). Results averaged across 100 runs (different random seed), 100K trials each.  $g_V = 0.005$  for dummy,  $g_V = 0$  for 1loc. Compare with Figure 6A in main text.



**Figure S5. SDE and learning generalization.**

Plotted are re-learning times (in trial units) as a function of bias, for three lags (1, 4, 10), with (top) and without (bottom) sticky priors. The results qualitatively demonstrate the phenomenon observed in the experiments: larger biases correlate with faster re-learning upon stimulus change. Color code represents mean volatility ( $Q$ ) corresponding to specific relearning-lag pairs.

- Falmagne J.-C., Cohen S. P., Dwivedi A. (1975) Two-choice reactions as an ordered memory scanning process. *Attention and Performance V*. pp. 296–344
- Fischer J., Whitney D. (2014) Serial dependence in visual perception. *Nature Neuroscience* **17**:738-743 <https://doi.org/10.1038/nn.3689> | PubMed
- Fornaciai M., Park J. (2018) Serial dependence in numerosity perception. *Journal of Vision* **18**:15 <https://doi.org/10.1167/18.9.15> | PubMed
- Fritsche M., Mostert P., de Lange F. P. (2017) Opposite effects of recent history on perception and decision. *Current Biology* **27**:590-595 <https://doi.org/10.1016/j.cub.2017.01.006> | PubMed
- Fritsche M., Spaak E., De Lange F. P. (2020) A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife* **9**:e55389 <https://doi.org/10.7554/eLife.55389> | PubMed
- Harris H., Gliksberg M., Sagi D. (2012) Generalized perceptual learning in the absence of sensory adaptation. *Current Biology* **22**:1813-1817 <https://doi.org/10.1016/j.cub.2012.07.059> | PubMed
- Harris H., Israeli D., Minshew N., Bonneh Y., Heeger D. J., Behrmann M., Sagi D. (2015) Perceptual learning in autism: over-specificity and possible remedies. *Nature Neuroscience* **18**:1574-1576 <https://doi.org/10.1038/nn.4129> | PubMed
- Harris H., Sagi D. (2018) Visual learning with reduced adaptation is eccentricity-specific. *Scientific Reports* **8**:608 <https://doi.org/10.1038/s41598-017-18824-7> | PubMed
- John-Saaltink E. S., Kok P., Lau H. C., De Lange F. P. (2016) Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *Journal of Neuroscience* **36**:6186-6192 <https://doi.org/10.1523/jneurosci.4390-15.2016> | PubMed
- Kalm K., Norris D. (2018) Visual recency bias is explained by a mixture model of internal representations. *Journal of Vision* **18**:1 <https://doi.org/10.1167/18.7.1> | PubMed
- 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**:4966-4970 <https://doi.org/10.1073/pnas.88.11.4966> | PubMed
- Karni a, Sagi D. (1993) The time course of learning a visual skill. *Nature* **365**:250-252 <https://doi.org/10.1038/365250a0> | PubMed
- Kersten D., Mamassian P., Yuille A. (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol* **55**:271-304 <https://doi.org/10.1146/annurev.psych.55.090902.142005> | PubMed
- Knill D. C., Pouget A. (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences* **27**:712-719 <https://doi.org/10.1016/j.tins.2004.10.007> | PubMed
- Kuai S.-G., Levi D., Kourtzi Z. (2013) Learning optimizes decision templates in the human visual cortex. *Current Biology* **23**:1799-1804 <https://doi.org/10.1016/j.cub.2013.07.052> | PubMed
- Lau W. K., Maus G. W. (2019) Visual serial dependence in an audiovisual stimulus. *Journal of Vision* **19**:20 <https://doi.org/10.1167/19.13.20> | PubMed
- Liberman A., Fischer J., Whitney D. (2014) Serial dependence in the perception of faces. *Current Biology* **24**:2569-2574 <https://doi.org/10.1016/j.cub.2014.09.025> | PubMed
- Lieder I., Adam V., Frenkel O., Jaffe-Dax S., Sahani M., Ahissar M. (2019) Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience* **22**:256-264 <https://doi.org/10.1038/s41593-018-0308-9> | PubMed
- Liu J., Lu Z.-L., Doshier B. A. (2012) Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision Research* **61**:15-24 <https://doi.org/10.1016/j.visres.2011.12.002> | PubMed
- Lu Z.-L., Doshier B. A. (2022) Current directions in visual perceptual learning. *Nature Reviews Psychology* **1**:654-668 <https://doi.org/10.1038/s44159-022-00107-2> | PubMed

- Manassi M., Kristjánsson Á., Whitney D. (2019) Serial dependence in a simulated clinical visual search task. *Scientific Reports* **9**:19937 <https://doi.org/10.1038/s41598-019-56315-z> | PubMed
- Manassi M., Liberman A., Kosovicheva A., Zhang K., Whitney D. (2018) Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review* **25**:2245-2253 <https://doi.org/10.3758/s13423-018-1454-5> | PubMed
- Manassi M., Murai Y., Whitney D. (2023) Serial dependence in visual perception: A meta-analysis and review. *Journal of Vision* **23**:18 <https://doi.org/10.1167/jov.23.8.18> | PubMed
- Manassi M., Whitney D. (2024) Continuity fields enhance visual perception through positive serial dependence. *Nature Reviews Psychology* **3**:352-366 <https://doi.org/10.1038/s44159-024-00297-x> | PubMed
- Mollon J. D., Danilova M. V. (1996) Three remarks on perceptual learning. *Spatial Vision* **10**:51-58 <https://doi.org/10.1163/156856896x00051> | PubMed
- Murai Y., Whitney D. (2021) Serial dependence revealed in history-dependent perceptual templates. *Current Biology* **31**:3185-3191 <https://doi.org/10.1016/j.cub.2021.05.006> | PubMed
- Ofen N., Moran A., Sagi D. (2007) Effects of trial repetition in texture discrimination. *Vision Research* **47**:1094-1102 <https://doi.org/10.1016/j.visres.2007.01.023> | PubMed
- Pascucci D., Mancuso G., Santandrea E., Della Libera C., Plomp G., Chelazzi L. (2019) Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLoS Biology* **17**:e3000144 <https://doi.org/10.1371/journal.pbio.3000144> | PubMed
- Petrov A. a, Doshier B. A., Lu Z.-L. (2005) The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review* **112**:715-743 <https://doi.org/10.1037/0033-295X.112.4.715> | PubMed
- Petrov A., Doshier B. A., Lu Z.-L. (2006) Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research* **46**:3177-3197 <https://doi.org/10.1016/j.visres.2006.03.022> | PubMed
- Piray P., Daw ND (2020) A simple model for learning in volatile environments. *PLoS Comput Biol* **16**:e1007963 <https://doi.org/10.1371/journal.pcbi.1007963> | PubMed
- Sagi D. (2011) Perceptual learning in Vision Research. *Vision Research* **51**:1552-1566 <https://doi.org/10.1016/j.visres.2010.10.019> | PubMed
- 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**:470-475 <https://doi.org/10.1038/nn.4490> | PubMed
- Suárez-Pinilla M., Seth A. K., Roseboom W. (2018) Serial dependence in the perception of visual variance. *Journal of Vision* **18**:4 <https://doi.org/10.1167/18.7.4> | PubMed
- Talluri B. C., Urai A. E., Tsetsos K., Usher M., Donner T. H. (2018) Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology* **28**:3128-3135 <https://doi.org/10.1016/j.cub.2018.07.052> | PubMed
- Tamaki M., Wang Z., Barnes-Diana T., Guo D., Berard A. V., Walsh E., Watanabe T., Sasaki Y. (2020) Complementary contributions of non-REM and REM sleep to visual learning. *Nature Neuroscience* **23**:1150-1156 <https://doi.org/10.1038/s41593-020-0666-y> | PubMed
- Tavoni G., Doi T., Pizzica C., Balasubramanian V., Gold J. I. (2022) Human inference reflects a normative balance of complexity and accuracy. *Nature Human Behaviour* **6**:1153-1168 <https://doi.org/10.1038/s41562-022-01357-z> | PubMed
- Urai A. E., De Gee J. W., Tsetsos K., Donner T. H. (2019) Choice history biases subsequent evidence accumulation. *eLife* **8**:e46331 <https://doi.org/10.7554/eLife.46331> | PubMed
- Van Bergen R. S., Jehee J. F. M. (2019) Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *Journal of Neuroscience* **39**:8164-8176 <https://doi.org/10.1523/jneurosci.3212-18.2019> | PubMed

Watanabe T., Sasaki Y. (2015) Perceptual learning: toward a comprehensive theory. *Annual Review of Psychology* **66**:197-221 <https://doi.org/10.1146/annurev-psych-010814-015214> | PubMed

Ying X. (2019) An overview of overfitting and its solutions. *Journal of Physics: Conference Series* **1168**:22022 <https://doi.org/10.1088/1742-6596/1168/2/022022>

Zhang T., Xiao L.-Q., Klein S. A., Levi D. M., Yu C. (2010) Decoupling location specificity from perceptual learning of orientation discrimination. *Vision Research* **50**:368-374 <https://doi.org/10.1016/j.visres.2009.08.024> | PubMed

## Peer reviews

### Reviewer #1 (Public review):

This paper presents a reanalysis of a large existing dataset to examine whether serial dependence effects-systematic influences of recent stimulus history on current perceptual judgments-are associated with generalization in perceptual learning. The central hypothesis is that extended, longer-range history effects (beyond the most recent trials) are beneficial for transfer across locations. The authors reanalyze data from a texture discrimination task in which observers discriminated peripheral target orientation against a line background, with performance quantified by stimulus-onset asynchrony thresholds. Three training conditions were compared: a fixed single-location condition, a two-location alternating condition, and a dummy-trial condition with frequent target-absent trials. Transfer was assessed after training at new locations. Serial dependence was quantified using history-sequence analyses and linear mixed-effects models estimating bias weights across stimulus lags, with summary measures distinguishing recent (1-3 trials back) and more distant (4-6 trials back) dependencies.

The authors report extended serial dependence effects, persisting up to 6-10 trials back, with substantial cumulative bias that remains stable across multiple days of training and is not correlated with overall performance thresholds. Recent history effects are stronger for faster responses, suggesting a contribution from decision- or response-related processes, whereas more distant effects decline within sessions, potentially reflecting adaptation dynamics. Critically, longer-range serial dependence is significantly stronger in training conditions that promote generalization than in the single-location condition. Individual differences in the strength and decay profile of distant history effects predict the magnitude of transfer across locations, whereas recent history effects do not. History effects are also correlated across trained locations, suggesting stable individual differences.

The authors interpret longer-range serial dependence as reflecting integrative processes that extract task-relevant structure over time, thereby supporting generalization, while shorter-range effects are attributed to more transient mechanisms such as priming or decision-level bias. The discussion connects these findings to Bayesian accounts of perceptual stability and to concepts of overfitting in machine learning.

The study offers a novel and thoughtful link between short-term serial dependence and long-term generalization in perceptual learning, helping bridge two literatures that are often treated separately. The large dataset enables robust estimation of individual differences, and the use of mixed-effects modeling appropriately accounts for variability across observers. The empirical distinction between recent and more distant history effects is well-supported and adds important nuance to interpretations of serial dependence. Converging evidence from both group-level comparisons and individual-level correlations strengthens the central conclusions.

Comments on revisions:

The authors have effectively addressed my concerns. The new robustness analyses (Supp. Fig. S3), supplementary toy model, clearer DDM-based mechanistic distinctions, and expanded discussion of limitations and generality fully resolve my original points.

<https://doi.org/10.7554/eLife.109830.2.sa2>

### Reviewer #3 (Public review):

Summary:

This reanalysis of a classic study of visual perceptual learning in a texture discrimination task convincingly demonstrates the presence of sequential dependence effects, commonly seen in response time analyses in 2-alternative tasks, on response accuracy in the texture task in visual periphery and in a simultaneous central letter report at fixation. Overall, this paper provides a new and interesting analysis of the effects of sequential dependencies from trial to trial on performance, learning, and generalizability in perceptual learning.

Strengths:

This new analysis of sequential dependency effects (SDEs) extends commonly observed sequential effects in two-choice reaction times to accuracy and relates them to response accuracy during visual learning in a frequently used perceptual learning task. The paper makes a convincing case that different conditions known to impact generalization of learning to a second visual location also expresses quantitatively distinct n-back SDEs.

Weaknesses:

Additional analyses now back up the analysis of effects of SDEs using trials selected to enhance the size of the effects, specifically when the current trial is low visibility and the prior trial is of high visibility. The authors now provide a practical analytic reason for this choice.

Comments on revisions:

The revision has successfully addressed comments in the original reviews.

<https://doi.org/10.7554/eLife.109830.2.sa1>

### Author response:

The following is the authors' response to the original reviews.

#### **Public Reviews:**

##### **Reviewer #1 (Public review):**

*This paper presents a reanalysis of a large existing dataset to examine whether serial dependence effects-systematic influences of recent stimulus history on current perceptual judgments-are associated with generalization in perceptual learning. The central hypothesis is that extended, longer-range history effects (beyond the most recent trials) are beneficial for transfer across locations. The authors re analyze data from a texture discrimination task in which observers discriminated peripheral target orientation against a line background, with performance quantified by stimulus-onset asynchrony thresholds. Three training conditions were compared: a fixed single location condition, a two-location alternating condition, and a dummy-trial condition with frequent target-absent trials. Transfer was assessed after training at new locations. Serial dependence was quantified using history-sequence analyses and linear mixed effects models*

*estimating bias weights across stimulus lags, with summary measures distinguishing recent (1-3 trials back) and more distant (4-6 trials back) dependencies.*

*The authors report extended serial dependence effects, persisting up to 6-10 trials back, with substantial cumulative bias that remains stable across multiple days of training and is not correlated with overall performance thresholds. Recent history effects are stronger for faster responses, suggesting a contribution from decision- or responderelated processes, whereas more distant effects decline within sessions, potentially reflecting adaptation dynamics. Critically, longer-range serial dependence is significantly stronger in training conditions that promote generalization than in the single-location condition. Individual differences in the strength and decay profile of distant history effects predict the magnitude of transfer across locations, whereas recent history effects do not. History effects are also correlated across trained locations, suggesting stable individual differences.*

*The authors interpret longer-range serial dependence as reflecting integrative processes that extract task-relevant structure over time, thereby supporting generalization, while shorter-range effects are attributed to more transient mechanisms such as priming or decision-level bias. The discussion connects these findings to Bayesian accounts of perceptual stability and to concepts of overfitting in machine learning.*

*The study offers a novel and thoughtful link between short-term serial dependence and long-term generalization in perceptual learning, helping bridge two literatures that are often treated separately. The large dataset enables robust estimation of individual differences, and the use of mixed-effects modeling appropriately accounts for variability across observers. The empirical distinction between recent and more distant history effects is well-supported and adds important nuance to interpretations of serial dependence. Converging evidence from both group-level comparisons and individuallevel correlations strengthens the central conclusions.*

*Several limitations should be addressed. First, the study relies entirely on previously collected data, without experimental manipulations designed to selectively isolate serial dependence mechanisms. Filtering choices, while theoretically motivated, may amplify history effects in ways that are difficult to quantify. Second, sequential dependencies can arise from multiple sources, including gradual updating of internal weight structures, adaptation processes, and history-dependent biases in decisionmaking. The current analyses do not clearly separate these contributions, limiting mechanistic attribution of long-range effects. Third, the conclusions are based on a single perceptual task, leaving open questions about generality across paradigms. Finally, while the discussion references computational ideas, no explicit modeling is provided to test whether plausible learning rules can jointly account for the observed history profiles and transfer effects.*

We now address these issues in the manuscript (see below for detailed responses) and provide a toy model (supplementary material) where the observed effects are explained by simple learning mechanisms.

*The findings align with theoretical frameworks that conceptualize perceptual learning as gradual reweighting of stable sensory representations at the decision stage (e.g., Petrov et al., 2005). Trial-by-trial updates in these models naturally give rise to sequential dependencies and sensitivity to training statistics. The observation that longer-range history effects predict generalization is consistent with broader temporal integration supporting more flexible learning, while narrower integration may lead to specificity. The results also indicate that multiple mechanisms - including decisionlevel biases and adaptation - may coexist with reweighting processes, highlighting the value of hybrid accounts.*

*In summary, this is a careful and data-rich reanalysis that highlights a potentially important role for serial dependence in enabling generalization during perceptual learning. While the underlying mechanisms remain underspecified, the evidence supporting the reported associations is strong, and the work provides a valuable empirical foundation for further experimental and modeling efforts.*

**Reviewer #2 (Public review):**

*This manuscript investigates how people's perceptual reports are influenced by events and trials in the past, and how this long-range dependence relates to broader learning across locations in a visual learning task. The authors present clear and internally consistent analyses showing that extended temporal integration is associated with greater generalization of learning. The study is thought-provoking and may contribute meaningfully to understanding how short-term influences and long-term improvement interact, although several interpretational points would benefit from clarification.*

**Strengths:**

*(1) The manuscript identifies unusually long-range perceptual biases extending up to ten trials back, which is a striking and potentially important finding.*

*(2) The association between strong long-range dependence and greater learning generalization is clearly documented and supported by consistent analyses.*

*(3) The dataset is large and rich, and the authors apply repeated and well-controlled analyses that give confidence in the stability of the effects.*

*(4) The writing is generally clear, and the manuscript raises interesting conceptual links between temporal integration and generalization of learning.*

**Weaknesses / Points Requiring Clarification:**

*(1) The manuscript repeatedly equates generalization with increased efficiency, but this relationship is not universally true. In some populations or tasks, excessive generalization can reduce task-specific efficiency. The authors should discuss this context-dependence to clarify when generalization is beneficial versus detrimental.*

We agree with the reviewer that generalization does not strictly imply increased efficiency; in some contexts, over-generalization can indeed be detrimental. We now explicitly note in the Introduction that serial dependence can impair performance when stimuli vary randomly across trials. We have reviewed the manuscript to ensure we do not explicitly equate generalization with efficiency. Our argument is specifically that long-range SDEs support the transfer of learning (generalization).

*(2) Serial dependence is also present, though smaller, in the central fixation task. It remains unclear whether this bias could contribute to the serial dependence observed in the main task. The authors should clarify whether the two biases are independent or whether the central-task bias might partially influence orientation judgments in the main task.*

These two tasks are independent, one requires T/L discrimination the other V/H discrimination. See our detailed response below.

*(3) Several figure captions and labels contain minor inconsistencies in formatting and terminology. Careful proofreading would improve clarity.*

We thank the reviewer for pointing this out and have proofread the captions to improve formatting and terminology consistency throughout.

**Reviewer #3 (Public review):**

*This reanalysis of a classic study of visual perceptual learning in a texture discrimination task convincingly demonstrates the presence of sequential dependence effects, commonly seen in response time analyses in 2-alternative tasks, on response accuracy in the texture task in the visual periphery and in a simultaneous central letter report at fixation. Overall, this paper provides a new and interesting analysis of the effects of sequential dependencies from trial to trial on performance, learning, and generalizability in perceptual learning.*

**Strengths:**

*This new analysis of sequential dependency effects (SDEs) extends commonly observed sequential effects in two-choice reaction times to accuracy and relates them to response accuracy during visual learning in a frequently used perceptual learning task. The paper makes a convincing case that different conditions known to impact generalization of learning to a second visual location also express quantitatively distinct n-back SDEs.*

**Weaknesses:**

*Most of the new analyses emphasize the effects of SDEs, including trials designed to enhance the size of the effects, specifically when the current trial is low visibility, and the prior trial is of high visibility. Unless there is an argument that learning and subsequent generalization primarily occur in low-visibility trials, the presentation should also include displays and an emphasized discussion of analysis for all trials, unfiltered.*

We analyze effects on close to threshold (small-medium SOA) current targets preceded by above threshold (high SOA) reference targets. This is motivated by both technical issues and theoretical assumptions. In psychophysics, when using percent correct as a measure of performance, bias cannot be reliably estimated at or near ceiling performance, as correct responses leave little room for bias to manifest. Regarding the ‘easy’ targets used as a reference, having them at low SOA introduces uncertainty as for the reference orientation against which bias is measured, with their perceptual effect being ambiguous. Theoretically, we note that in perceptual learning with threshold targets, the introduction of clear targets in the absence of feedback enables learning (see Discussion, where we added: ‘Most interestingly, in our experiments without feedback on the texture task, the experimental conditions yielding the strongest bias were also found to enhance learning in the absence of feedback (Liu et al., 2012)’)

We have addressed this concern also by conducting additional robustness analyses with unfiltered prior-trial history. We analyzed data without the prior-visibility filter; results are presented in a new Supplementary Figure S3 and confirm our main findings (see addition to Methods: “Finally, to verify that our findings are not artifacts of these filtering choices, we also conducted control analyses including all prior-trial history regardless of visibility; these results are presented in Supplementary Figure S3 and confirm the robustness of our main findings.”).

**Recommendations for the authors:**

**Reviewer #1 (Recommendations for the authors):**

*(1) How manipulations of stimulus statistics, uncertainty, or feedback could selectively engage different forms of serial dependence*

We expect serial dependence to be modulated by all these parameters. In classical SDT, stimulus statistics are known to affect response bias, as are temporal correlations in stimulation sequences. We note in the manuscript that we employed random sequences (50% chance for V and 50% for H targets), eliminating expectation-based biases toward either orientation. Stimulus uncertainty is known to increase serial dependence, as we also found here. Feedback is also expected to have an effect, the literature is somewhat ambiguous about this, but this may also depend on experimental design. We note that the main task studied here (TDT) had no feedback while the central T/L task did have feedback, both showing serial dependencies. In the manuscript we point to reviews of SDE where much of this is discussed.

*(2) How explicit computational models could help distinguish decision bias from structural learning*

We use the drift diffusion model (DDM) to distinguish decision bias (starting point in DDM) from structural learning (changes in drift rate). DDM predicts that decision bias is short lived, mainly affects fast reaction times (RT) while biases due to drift rate asymmetry persists to long RTs. We present these results in Figure 3.

*(3) Whether similar relationships are observed in other perceptual domains*

We are not aware of any other study linking serial dependence and perceptual learning or reporting such a link. We expect the link between long-range serial dependence and learning generalization to extend beyond the TDT (see new paragraph in Discussion). We hope this framework will motivate similar analysis in other labs where comparable datasets exist.

*(4) How sensitive are the results to the filtering choices used in the analysis?*

We analyze effects on close to threshold (small-medium SOA) current targets preceded by above threshold (high SOA) reference targets. This is motivated by both technical issues and theoretical assumptions. In psychophysics, when using percent correct as a measure of performance, bias cannot be reliably estimated at or near ceiling performance, as correct responses leave little room for bias to manifest. Regarding the 'easy' targets used as a reference, having them at low SOA introduces uncertainty as for the reference orientation against which bias is measured, with their perceptual effect being ambiguous. Theoretically, we note that in perceptual learning with threshold targets, the introduction of clear targets in the absence of feedback enables learning (see Discussion, where we added: 'Most interestingly, in our experiments without feedback on the texture task, the experimental conditions yielding the strongest bias were also found to enhance learning in the absence of feedback (Liu et al., 2012)')

We have addressed this concern also by conducting additional robustness analyses with unfiltered prior-trial history. We analyzed data without the prior-visibility filter; results are presented in a new Supplementary Figure S3 and confirm our main findings (see addition to Methods: "Finally, to verify that our findings are not artifacts of these filtering choices, we also conducted control analyses including all prior-trial history regardless of visibility; these results are presented in Supplementary Figure S3 and confirm the robustness of our main findings.").

**Reviewer #2 (Recommendations for the authors):**

*(1) Clarify mechanisms underlying long-range serial dependence. Please better distinguish possible sources of serial dependence (e.g., decision bias, adaptation, reweighting) and clarify which interpretations are supported or remain ambiguous given the current analyses*

Our manuscript discusses the mechanisms underlying the dissociation between recent and distant SDEs in the Discussion section. Specifically, we report that:

Recent SDEs are RT-dependent (stronger with faster responses) consistent with decision-level criterion shifts (Dekel & Sagi, 2020)

Distant SDEs are RT-independent consistent with neural reweighting/template updating

We also discuss the role of sensory adaptation in truncating long-range integration, supported by within-session decline of SDEs, reduced distant SDEs in the 1loc condition, and the original findings by Harris et al. (2012).

We have added an explicit acknowledgment that our correlational approach cannot definitively establish causality (see addition to Discussion: "While these converging findings support distinct mechanisms for recent and distant SDEs, our correlational approach cannot definitively establish causality, and targeted experimental manipulations would further strengthen these interpretations.").

| (2) *Test robustness to analytic choices*

We have conducted robustness analyses by removing the prior-trial visibility filter. The results are presented in a new Supplementary Figure S3 and confirm that our key findings remain qualitatively unchanged (see addition to Methods referencing Supplementary Figure S3).

| (3) *Strengthen the computational link*

We have expanded the Discussion to reference relevant computational models and specify predictions for future modeling work. We now cite Petrov et al. (2005). We provide a toy model implementing trial-by-trial template update that show SDE that is correlated with learning transfer. Importantly, in this model, long range SDE is a consequence of learning dynamics (see new paragraph in Discussion, and model simulation in supplementary material).

| (4) *Discuss generality and experimental tests. Briefly address whether similar effects are expected across other tasks or sensory domains, and outline experimental manipulations that could causally test the role of serial dependence in generalization.*

We have added discussion of generality across perceptual domains and outlined the prediction that future work could test the SDE-generalization link in other tasks where both phenomena have been documented (see new paragraph in Discussion).

| *Reviewer #2 (Public Review - Point 2): Central task SDE independence*

The SDEs observed in the central letter task and peripheral TDT are likely independent, as they involve different stimulus features (letter identity vs. orientation), different response mappings, and show distinct performance patterns across conditions. The absence of condition differences in central-task SDEs (described in the Results section under "SDE differences between conditions" end of paragraph), despite robust differences in TDT SDEs, further suggests that the peripheral orientation biases are not contaminated by central-task response tendencies. Note that the central task was fixed across conditions, stayed at fixation when location was changed, and when dummy trials were presented.

| **Reviewer #3 (Recommendations for the authors):**

| (1) *Reference to Falmagne, Cohen, & Dwivedi (1975)*

We have added this reference to the Introduction, acknowledging the historical foundation of sequential effects in perceptual decisions

*(2) The SDE data of Figure 1 are (per the figure legend) from the 1 loc data of Harris et al., "pooled over all testing days", and filtered for trials with low-visibility current targets ( $SOA < SOA\text{-threshold} + 20\text{ms}$ ). Specify whether this threshold criterion is on a per-subject basis. State in the legend that "all testing days" includes Days 1-8 (4 days with the first location and another 4 days testing generalization to a second location).*

We have revised the Figure 1 legend to clarify:

"Days 1–8; 4 days at the first location and 4 days at the second location to assess generalization"

"calculated on a per-subject basis"

*(3) The leadup emphasizes that the analysis in the figure emphasizes trials where the effect is expected to be as large as possible (cited as  $40 \pm 3\%$ ), while visible current targets (at  $n$ ) biases were  $5 \pm 1\%$ .*

See below, after (4).

*(4) Unless a theoretical position associates learning just with low visibility (if so, explain), consider including two other panels showing the sequential dependencies for all trials, and the linear model weights over the last 10 trials for all trials.*

We acknowledge that the main analyses emphasize conditions that maximize SDE expression. To verify robustness, we conducted control analyses including all prior-trial history regardless of visibility; these results are presented in Supplementary Figure S3 and confirm our main findings.

There are both theoretical and technical justifications for the filtering applied:

It is well known that learning, in particular without feedback (as in our TDT), is facilitated by a mixture of threshold level stimuli and suprathreshold easy trials (e.g., Liu et al., 2012).

Technically, it is impossible to measure bias with highly discriminable stimuli where performance is perfect or close to it, thus such trials are expected to dilute the measured effect. On the other hand, when considering serial effects from low sensitivity trials, we face an uncertainty involved in defining the actual orientation relative to which the bias needs to be computed.

*(5) Figure S1 seems to indicate that average thresholds over all days (location 1 and location 2) are unrelated to the sequential dependence across subjects and that the amount of learning in location 1 is unrelated to the sequential dependencies across subjects in all the varied conditions. Since Figure S1 includes all 50 subjects, it includes some conditions with dummy trials interspersed. Clarify in the description whether the dummy trials are ignored for the purposes of the SDE analyses.*

We have clarified in the Methods how trials are handled in the analysis: "To preserve the precise temporal structure of the data, all trials were included in the sequential n-back count across all experimental conditions, thus dummy trials were counted as time bins but their contribution was ignored. In the Linear Mixed Effects (LME) analysis, we modeled these trial types using distinct regressors: each n-back lag included separate predictors for visible and invisible targets, further differentiated by trial type (dummy vs. target) and relative location (ipsilateral vs. contralateral) where applicable. The SDE values reported here reflect only the influence of relevant target-present history trials; the effects of other history types (e.g.,

dummy trials), while estimated to ensure the temporal integrity of the model, are not presented."

*(6) The conclusion from this analysis seems to be that the overall average threshold and the amount of initial learning are both uncorrelated with the strength of sequential dependencies across subjects. This conclusion should be added to the description in the main paper.*

This finding is now discussed in the Discussion section, referring to the main Results section [No significant correlation was found between biases and SOA thresholds across observers ( $r = -0.13$ ,  $p = 0.37$ , average across days 1-8), nor between biases and improvements in performance at the first location ( $r = -0.09$ ,  $p = 0.54$ , average across days 1-4), suggesting that the magnitude of serial dependence does not predict the overall amount of perceptual learning (Supplementary Figure S1)].

*(7) Decay of SDE section clarifications*

We have made the following clarifications:

RT definition: Added to Methods: "The reaction time (RT) used in the analysis was defined as RT(TDT) – RT (fixation task), where RT for each task was measured from stimulus onset."

N-back counting: Clarified in Methods (see response to point 5 above): all trials were included in the chronological sequence; the LME analysis assigned separate predictors at each lag for visible/invisible targets and for trial categories (dummy vs. target) and locations (ipsilateral vs. contralateral). The results reported do not include effects of dummy trial, except where response dependent SDE was reported (Fig 2a, SDE for response key).

2loc n-back effect: The longer-range effects in the 2loc condition likely reflect reduced adaptation allowing longer temporal integration, combined with the location-selective nature of SDEs.

RT and mechanism interpretation: The manuscript discusses that the critical observation is the qualitative difference in RT sensitivity between recent and distant SDEs, consistent with the drift-diffusion framework where criterion shifts are RT-dependent while drift bias is RT-independent (Dekel & Sagi, 2020). We have added an acknowledgment of the correlational limitations of this interpretation.

Moving figures to supplement: We prefer to keep Figures 4 and 5 in the main text as they document important dynamics supporting our mechanistic interpretation.

<https://doi.org/10.7554/eLife.109830.2.sa0>