

<https://doi.org/10.1038/s41539-025-00312-7>

Anterograde interference in multitask perceptual learning



Jia Yang^{1,2,3,4}, Fang-Fang Yan^{1,2}, Tingting Wang^{1,2}, Zile Wang^{1,2}, Qingshang Ma^{1,2}, Jinmei Xiao^{1,2}, Xianyuan Yang^{1,2}, Zhong-Lin Lu^{5,6,7}✉ & Chang-Bing Huang^{1,2}✉

Learning to perform multiple tasks robustly is a crucial facet of human intelligence, yet its mechanisms remain elusive. Here, we formulated four hypotheses concerning task interactions and investigated them by analyzing training sequence effects through a continual learning framework. Forty-nine subjects learned seven tasks sequentially, each of the seven groups following a distinct sequence. Results showed that subjects learning a task later in a sequence exhibited poorer performance in six tasks (Contrast, Vernier, Face, Motion, Auditory, and *N*-back tasks, except for the Shape task) compared to those who learned this task earlier. Interestingly, sequence position had minimal impact on forgetting. A complementary dual-task experiment corroborated these findings. Through detailed analyses of session and block learning curves, we revealed task-specific anterograde interference, but no retrograde interference. These findings support the integrated reweighting theory and shed light on the meta-plasticity mechanism governing how human brain balances plasticity and stability.

Humans are not only extremely sensitive in performing a wide range of simple perceptual tasks but also exceptionally proficient in performing complicated tasks comprised of multiple perceptual processes¹. While some of our perceptual skills develop during childhood, others are acquired and/or refined through perceptual learning well into adulthood^{2–4}. This “natural” perceptual expertise underscores our capacity to master multiple perceptual tasks through perceptual learning. Studies on perceptual expertise in specialized domain^{5–7} and perceptual rehabilitation^{8,9} also highlight the importance of training across multiple perceptual tasks to achieve proficiency in the real world. Despite strong evidence of learning-induced plasticity across all perceptual and cognitive domains¹⁰, the brain’s ability to reconcile the need for plasticity with the maintenance of existing expertise remains a puzzle¹¹. In this study, we investigate interactions across multiple perceptual tasks learned through sequential training.

We consider four hypotheses regarding task interactions: independence, facilitation, retrograde interference, and anterograde interference (Fig. 1). To test these hypotheses, we would employ multiple perceptual tasks and explicitly manipulate their sequence position in sequential training. The simplest design would involve two tasks in two sequences: In the first sequence, task A is trained first, followed by training in task B and testing in task A; and in the second sequence, task B is trained first, followed by training in task A and testing in task B. If the two tasks are completely independent^{12–15}, the sequence

of training would have no impact on any performance measures of either task (Fig. 1a). If training facilitates learning, tasks trained later in the sequence would benefit from the training of tasks earlier in the sequence, resulting in “learning to learn”--- improved initial performance and/or faster learning^{16–18} (Fig. 1b). Retrograde interference^{19,20}, also known as “catastrophic forgetting” in artificial intelligence^{21,22}, suggests that learning a new task would disrupt performance in a previously learned task (Fig. 1c). Lastly, anterograde interference²³ is defined as the detrimental effect of prior learning on subsequent learning in a new task, manifested as reduced initial performance, a slower learning rate, or both, while performance in the first-trained task remains preserved after learning the second task (Fig. 1d).

Although there is evidence for all four types of interactions, most studies have not used the complete design to fully test these hypotheses. The closest is a study by Fahle¹³ who trained three hyperacuity tasks sequentially in all six possible training sequences. He found that the average performance across the three tasks was slightly better in the first sequence position than the other sequence positions and conclude that the tasks were mostly independent, with a trend of anterograde interference. Fahle and Morgan¹² trained three-dot Vernier and bisection tasks sequentially in two sequences and found evidence for anterograde interference. However, because there was no re-test of the first trained task following the completion of the second task, we do not know whether there was retrograde interference. Using

¹Institute of Psychology, Chinese Academy of Sciences, Beijing, China. ²Department of Psychology, University of Chinese Academy of Sciences, Beijing, China.

³School of Psychological and Cognitive Sciences, Peking University, Beijing, China. ⁴Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China.

⁵Division of Arts and Sciences, New York University Shanghai, Shanghai, China. ⁶Center for Neural Science and Department of Psychology, New York University, New York, NY, USA. ⁷NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai, China. ✉e-mail: zhonglin@nyu.edu;

huangcb@psych.ac.cn

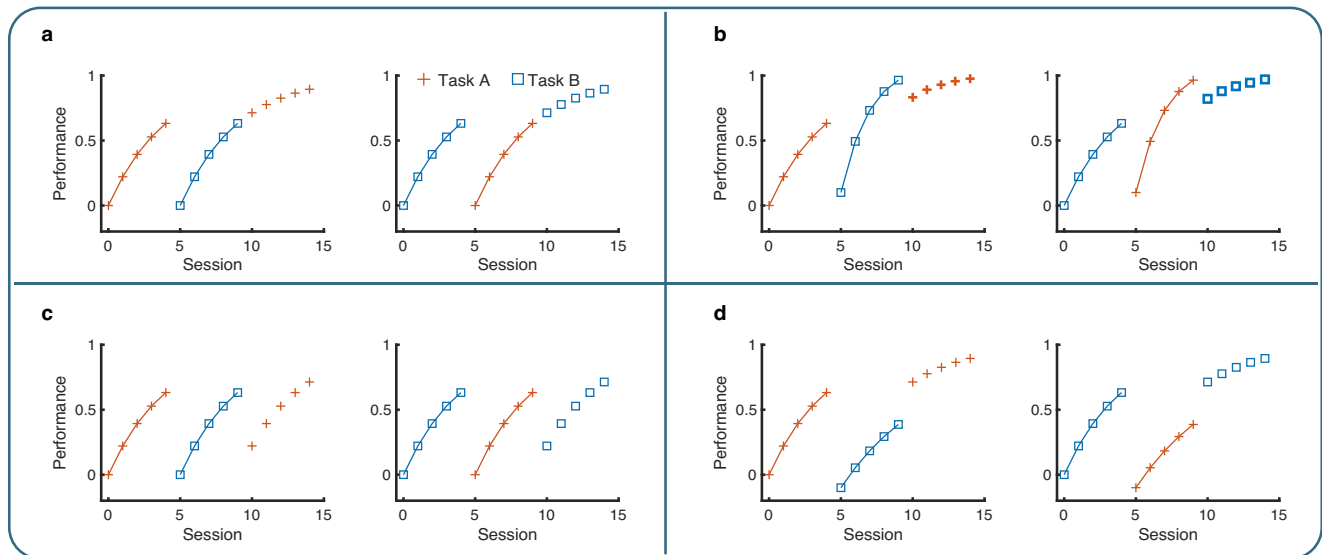


Fig. 1 | Four hypotheses on task interactions in sequential training of two tasks. **a** Independence: No effect of training sequence. **b** Facilitation: Performance in the task trained later benefits from training of the earlier task. **c** Retrograde Interference: Training of a new task disrupts the performance of a learned task. **d** Anterograde Interference: Training of a task hinders subsequent training of a new task.

Training of a new task disrupts the performance of a learned task. **d** Anterograde Interference: Training of a task hinders subsequent training of a new task.

similar tasks with alternating blocks of the two training tasks in two sequences, Huang et al.¹⁵ found that learning of the two tasks was independent, with no evidence for any interaction.

Other studies employed two tasks, designating one as the training task and the other as the transfer task in one training sequence, without explicitly manipulating the order of the two tasks. These studies found that performing the training task either enhanced the initial performance^{24–27}, sped up the learning rate²⁸, or had no effect on the performance^{24,26} of the transfer task, and overlearning would induce anterograde interference¹⁹. Additional tests on the training task after the completion of the transfer task in some of these studies found that there was no long-term retrograde interference^{24,25}, and short-term retrograde interference can be avoided by overlearning¹⁹ or extended time interval between two training phases²⁰. One study, which sequentially trained five tasks with a common task structure, demonstrated that learning rate increased as training experience accumulated¹⁷. While these studies have revealed some important effects of the training task on the transfer task, the lack of explicit manipulation of the training sequence of the trained and transfer tasks renders them inadequate for fully evaluating the four hypotheses outlined above. For example, although studies have found that playing action video game could facilitate the performance and learning in perceptual and cognitive tasks^{18,28}, it remains unclear whether training in these perceptual and cognitive tasks can, in turn, enhance performance in video game playing.

The more common paradigm trained only one task but included pre- and post-training tests on additional untrained tasks, while manipulating the designation of the trained and untrained tasks across groups. These studies found that the training on one task either had no impact on²⁹, or benefited the performance of^{30–34} the untrained tasks. However, because they only trained one task, these studies can only uncover the impact of the trained task on the initial performance of the untrained tasks instead of any changes of learning dynamics.

The current study employed the learning of multiple tasks and explicitly manipulated their sequence position during training to investigate task interactions. In Experiment 1A, 49 subjects were trained in seven tasks (Fig. 2), and were evenly divided into seven groups, each receiving training in a distinct sequence. Additionally, a subset of them also participated in a re-test of these seven tasks several months later (Experiment 1B). In Experiment 2, fourteen subjects were trained in Vernier offset discrimination and visual shape search tasks, with half of them learning one task first,

and the other half learning the other task first. The dataset used in Experiment 1A includes seven different learning tasks spanning a broad range of perceptual domains to maximize variability, including low-level (e.g., contrast detection), mid-level (e.g., visual shape search), and high-level tasks (e.g., face view discrimination), as well as tasks across different modalities (visual, auditory, and audiovisual working memory, such as the *N*-back task). This diverse task set better approximates real-world learning and allows us to explore whether tasks of varying complexity and modality share common or task-specific learning mechanisms, and how they interact. When we originally collected this data, our goals were to examine (1) individual differences, (2) the functional forms of learning curves, and (3) the effects of training sequences in perceptual learning. Experiment 2 was conducted to validate the key findings in Experiment 1 using a simpler experimental design. Papers on individual differences³⁵ and the functional forms of learning curves³⁶ have already been published. In this study, we implemented session-by-session Linear Mixed-Effects Models (LMMs) and built a comprehensive block-by-block model to assess the influence of training sequence position on learning and long-term forgetting processes.

Results

Effects of sequence position in learning seven tasks (Experiment 1A)

With a total of forty-nine subjects (22 males, 23.4 ± 2.5 years) trained in seven tasks across seven distinct training sequences (Fig. 2a, b), Experiment 1A employed a two-factor mixed design. Task served as a within-subject factor, while training sequence acted as a between-subject factor (See “Methods” for detailed information). Each specific task, such as motion direction discrimination highlighted in red in Fig. 2c, was learned in each of the seven sequence positions across the seven distinct training sequence groups.

To visualize the influence of sequence position on performance, we calculated the average session-by-session learning curves of subjects who received training in sequence positions 1, 2, and 3 (“early training”), and those who received training in sequence positions 5, 6, and 7 (“late training”) for each task. The results (Fig. 3) showed that, for five out of the seven tasks (Motion, Vernier, Contrast, Face, and Auditory), subjects who underwent training in the early sequence position exhibited better performance than those who received training later for the same task. For the other two tasks (Shape and *N*-back), the training sequence position had no discernible effect.

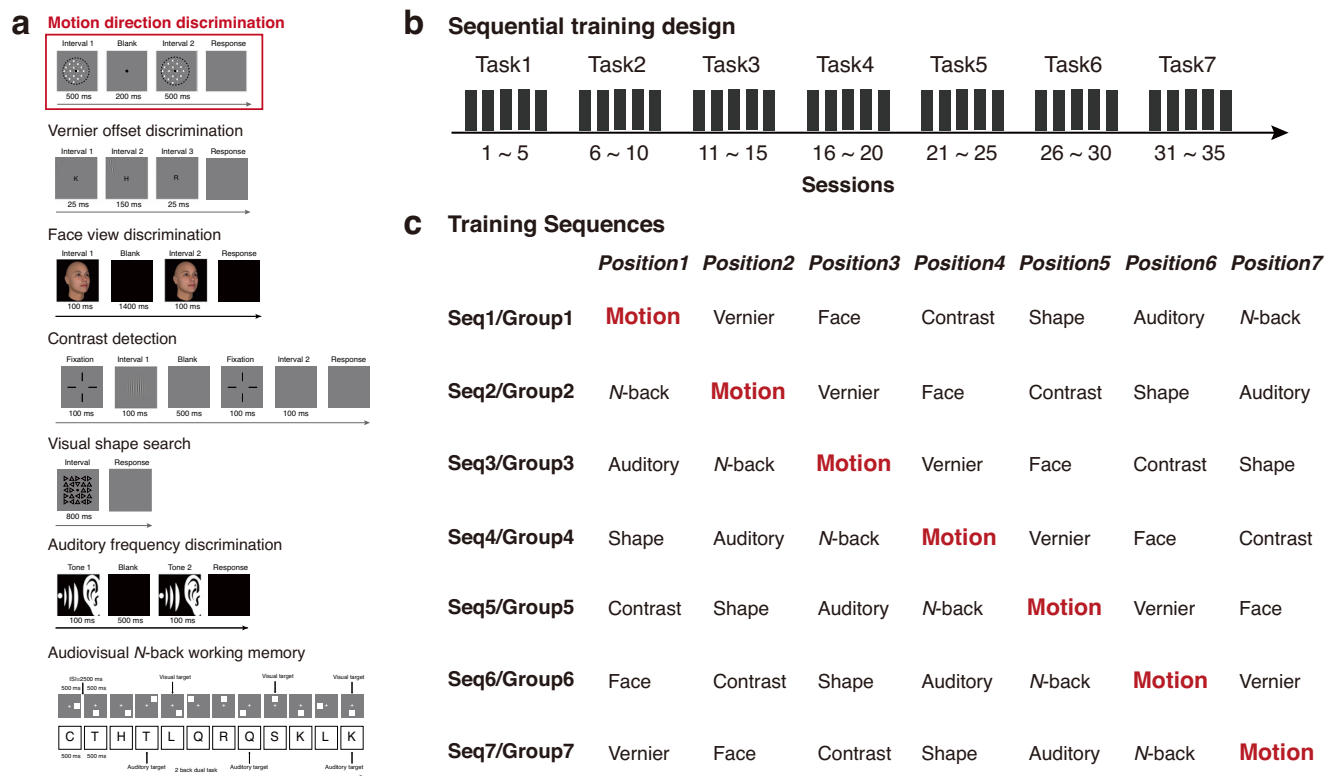


Fig. 2 | Design and Procedures (Experiment 1A). **a** Illustrations of the procedures of the seven training tasks. The facial images used in the face view discrimination task were generated using FaceGen software and are entirely synthetic, which do not

depict any real individuals. **b** Illustration of the sequential training design. **c** The seven distinct training sequences. The motion direction discrimination task was highlighted to depict the seven training sequence positions for each task.

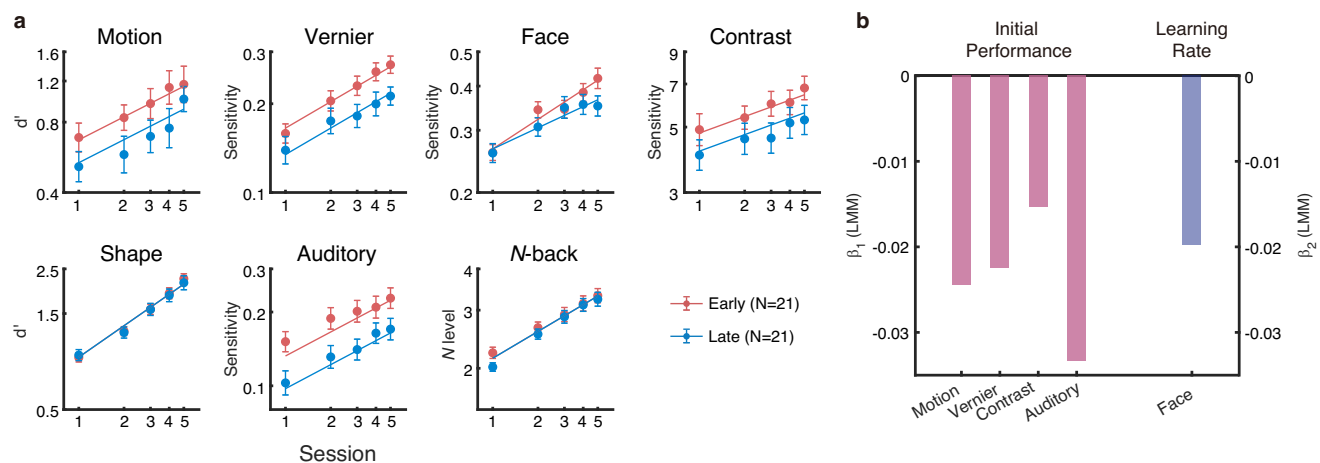


Fig. 3 | Effects of training sequence position on learning seven tasks (Experiment 1A, $N = 49$). **a** The averaged session-by-session performances of subjects who received training in sequence positions 1, 2, and 3 (“early training”, represented by red dots), and subjects who received training in sequence positions 5, 6, and 7 (“late training”, represented by blue dots) for each of the seven tasks, along with the

predicted learning curves of the best-fitting LMM model (red and blue solid lines). Error bar denotes standard error. **b** Coefficients of sequence position on the initial performance (pink color) and the general learning rate (lavender color) for the five of seven tasks that exhibited significant sequence position effects.

To test these observations statistically, we employed LMMs to analyze the effect of sequence position in each task. Specifically, we compared the fits of four models to the seven average learning curves of each task, one for each sequence position. Modeling each learning curve as a linear function with an intercept (initial performance) and a slope (general learning rate) acting on log session number. M1 (2 parameters) assumes no effect of sequence position on both initial performance and learning rate; M2 (3 parameters) includes modulation (β_1) of initial performance by sequence position; M3 (3

parameters) includes modulation (β_2) of general learning rate by sequence position. M4 (4 parameters), the most saturated model, accounts for both modulations (β_1 and β_2). Model comparisons identified the best-fitting model for each task as the one statistically equivalent to M4, outperforming its reduced versions, and having the fewest parameters (see Methods).

Figure 3 only shows the averaged predicted learning curves from the best fitting model for the early-training (sequence positions 1, 2, and 3) and the late-training (sequence positions 5, 6, and 7) conditions. For the motion

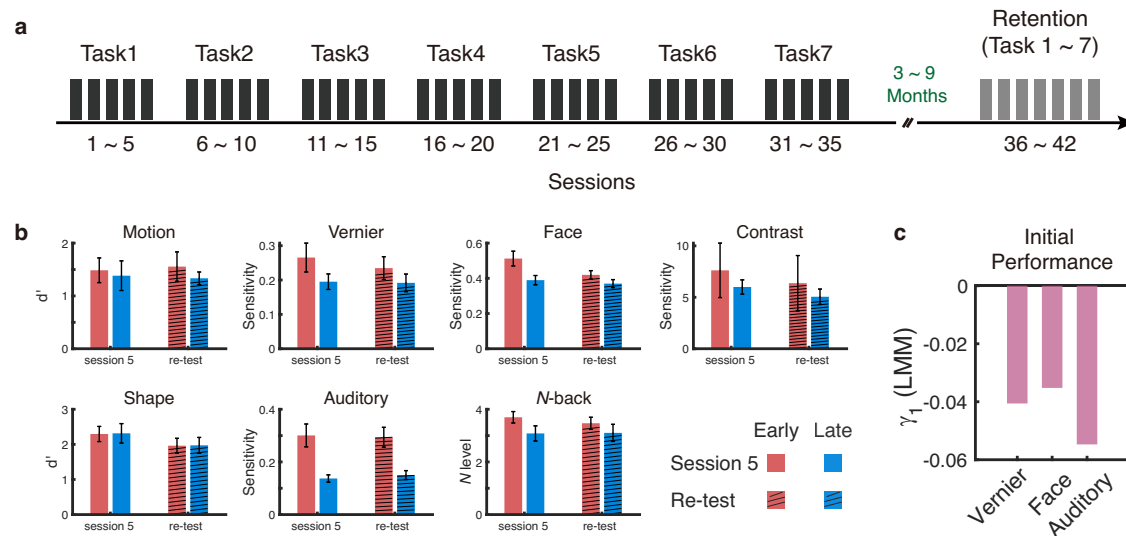


Fig. 4 | Design and results (Experiment 1B, $N = 17$). **a** Illustration of the sequential training design with a re-test. **b** Average performance of the subjects who received early training (red bar) and late training (blue bar) in their last training session

(Session 5) and re-test for each of the seven tasks. **c** The coefficients of sequence position on the initial performance in three of the seven tasks of the best fitting model.

direction discrimination, Vernier offset discrimination, contrast detection, and auditory frequency discrimination tasks, the best fitting model was M2 (Fig. 3a and Supplementary Table 1). In these tasks, M2 significantly outperformed M1 (Motion: M2 vs M1, $LR(1) = 6.35$, $p = 0.01$, 95% CI = [0.007, 0.022]; Vernier: M2 vs M1, $LR(1) = 19.64$, $p < 0.001$, 95% CI = [2.53×10^{-5} , 0.006]; Contrast: M2 vs M1, $LR(1) = 4.66$, $p = 0.03$, 95% CI = [0.018, 0.039]; Auditory: M2 vs M1, $LR(1) = 24.74$, $p < 0.001$, 95% CI = [2.53×10^{-5} , 0.006]), and was statically equivalent to M4 (Motion: M2 vs M4, $LR(1) = 0.002$, $p = 0.97$, 95% CI = [0.96, 0.98]; Vernier: M2 vs M4, $LR(1) = 0.46$, $p = 0.52$, 95% CI = [0.49, 0.55]; Contrast: M2 vs M4, $LR(1) = 0.05$, $p = 0.81$, 95% CI = [0.78, 0.83]; Auditory: M2 vs M4, $LR(1) = 0.09$, $p = 0.76$, 95% CI = [0.73, 0.79]). The negative coefficients of sequence position on initial performance (Fig. 3b, Motion: $\beta_1 = -0.024$, $p = 0.01$, 95% CI = [-0.04, -0.005]; Vernier: $\beta_1 = -0.022$, $p < 0.001$, 95% CI = [-0.03, -0.01]; Contrast: $\beta_1 = -0.015$, $p = 0.03$, 95% CI = [-0.03, -0.001]; Auditory: $\beta_1 = -0.033$, $p < 0.001$, 95% CI = [-0.05, -0.02]) indicate that subjects who underwent later training exhibited worse initial performance in those tasks.

For the face view discrimination task, the best fitting model was M3 (M3 vs M1, $LR(1) = 6.45$, $p = 0.01$, 95% CI = [0.007, 0.022]; M3 vs M4, $LR(1) = 0.27$, $p = 0.56$, 95% CI = [0.53, 0.60]), with a negative coefficient of sequence position on general learning rate ($\beta_2 = -0.020$, $p = 0.01$, 95% CI = [-0.035, -0.004]), indicating slower general learning rates for participants receiving training later.

For the shape search and N-back tasks, M1 was the best fitting model, indicating no significant effect of sequence position (Shape: M4 vs M1, $LR(2) = 2.77$, $p = 0.26$, 95% CI = [0.23, 0.29]; N-back: M4 vs M1, $LR(2) = 2.47$, $p = 0.30$, 95% CI = [0.27, 0.33]).

In summary, these results demonstrate that, for low-level and mid-level visual and auditory perceptual learning tasks, there was significant anterograde interference (Fig. 1d). Furthermore, the performance in high-level shape search and N-back working memory tasks was unaffected by sequence position.

Testing for retrograde interference (Experiment 1B)

To examine whether learning new tasks induces retrograde interference on previously acquired tasks, Experiment 1B re-evaluated a subset of 18 subjects from Experiment 1A after 3 to 9 months (Mean \pm SE: 6.67 \pm 1.78 months) following the completion of the initial experiment (Fig. 4a). The analysis in the main text only included 17 subjects, excluding one due to the estimated threshold in the auditory discrimination task being more than

5 standard deviations lower than the average performance of all the other subjects in the last training session of the task (Excluding the same subject from the analysis in Experiment 1A had only a very small effect on the results; see Supplementary Table 2 and Supplementary Fig. 1.)

Similar to Experiment 1A, the average performance of subjects who underwent early (sequence positions 1, 2, and 3) and late training (sequence positions 5, 6, and 7) was calculated for their last training and re-test sessions in each of the seven tasks (Fig. 4b). If retrograde interference had occurred, one would expect a greater reduction in performance for the early group compared to the late group when comparing their performance in the last training session and re-test. However, contrary to this expectation, the reduction in performance between the last training session and re-test was comparable between the early and late groups (all $p > 0.05$).

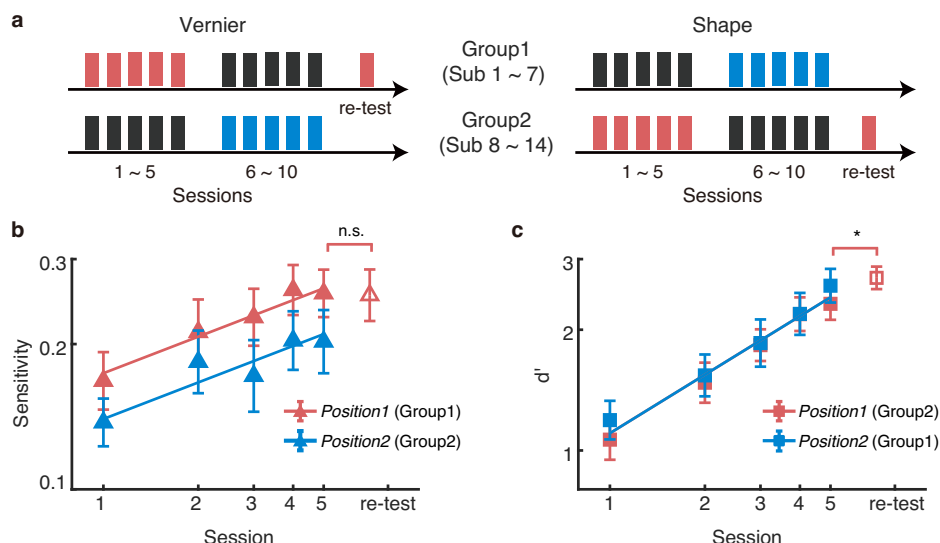
We employed four LMMs to assess the influence of sequence position on performance in both the last training and re-test sessions, considering all sequence positions for each task. N1 represented the most reduced model without any sequence position effect. N2 considered modulation (γ_1) of performance in the last training session by sequence position, while N3 considered modulation (γ_2) of the amount of performance reduction between the last training session and re-test by sequence position. Finally, N4 considered both the modulations of performance in the last training session (γ_1) and performance reduction between the last training session and re-test (γ_2) by sequence position (see Methods).

The results indicated that, for this subset of 17 observers, N2 was the best fitting model for the auditory frequency discrimination, Vernier offset discrimination, and face view discrimination tasks (Fig. 4c and Supplementary Table 3), suggesting a negative (i.e., anterograde) influence of training sequence position on performance in the last training session for these tasks (Vernier: $\gamma_1 = -0.04$, $t(31) = -2.81$, $p = 0.008$, 95% CI = [-0.07, -0.01]; Auditory: $\gamma_1 = -0.05$, $t(31) = -4.10$, $p < 0.001$, 95% CI = [-0.08, -0.03]; Face: $\gamma_1 = -0.04$, $t(31) = -4.15$, $p < 0.001$, 95% CI = [-0.05, -0.02]). For the remaining four tasks, the best fitting model was N1, indicating no significant effect of sequence position. Notably, in none of these tasks did sequence position have any significant effect on performance reduction between the last training session and re-test, as evidenced by selecting N3 as the best-fitting model. These results suggested that there was no retrograde interference in learning these seven tasks.

We also conducted the same analysis on the data from all 18 subjects, which changed the best-fitting model for the auditory frequency discrimination task to N1, while keeping the best-fitting models for the other tasks the same (Supplementary Table 4 and Supplementary Fig. 2).

Fig. 5 | Design and results (Experiment 2, $N = 14$).

a Sequential training of the Vernier and shape tasks, with a re-test of the first trained task conducted on the day after completing five training sessions of the second task. The red bars denote five training sessions and re-test sessions for the specific task (Left: Vernier; Right: Shape) trained in the first position while blue bars denote the same task trained in the second position. Black bars denote the five training sessions for the other task (Left: Shape; Right: Vernier). **b** The averaged performance of the two groups across five training sessions (solid triangles), and the averaged re-test performance of Group 1 (open triangle) in the Vernier task. **c** The averaged performance of the two groups across five training sessions (solid squares), and the averaged re-test performance of Group 1 (open square) in the Shape task.



Effects of sequence position in learning two tasks (Experiment 2)

Building on the findings from Experiments 1A and 1B, Experiment 2 aimed to replicate the observed effects in a simpler design. Fourteen subjects (7 males, 20.5 ± 2.5 years) participated, with half of them (Group1, S1–7) receiving training in the Vernier offset discrimination task first, while the other half (Group2, S8–14) started with the visual shape search task. Subsequently, all subjects received a re-test of the first trained task on the day following 5 sessions of training for the second task (Fig. 5a). The learning curves for the two groups in the two tasks, along with the re-test performance of the group that underwent initial training in each task, are depicted in Fig. 5b and c. To analyze this data, we employed the same four LMMs used in Experiment 1A.

For the Vernier offset discrimination task, M2 significantly outperformed M1 ($M2$ vs $M1$, $LR(1) = 6.66$, $p = 0.01$, 95% CI = [0.006, 0.02]) and was statically equivalent to M4 ($M2$ vs $M4$, $LR(1) = 0.08$, $p = 0.78$, 95% CI = [0.75, 0.80]). The estimated coefficient of sequence position on initial performance (β_1) was -0.09 , indicating a negative influence of sequence position on initial performance. Moreover, no significant performance reduction was observed in Group 1 during re-test ($t(6) = 0.002$, $p = 0.99$, Cohen's $d = 0.002$). For the visual shape search task, M1 was statistically equivalent to M4 ($M1$ vs $M4$, $LR(2) = 0.76$, $p = 0.68$, 95% CI = [0.69, 0.74]), signifying that sequence position had no significant impact on the shape search task, consistent with Experiment 1A. Additionally, improved performance was observed in Group 2 in re-test ($t(6) = -2.65$, $p = 0.04$, Cohen's $d = 0.66$), suggesting nonsaturated learning during the five sessions of training. In summary, anterograde interference was evident in the Vernier task, while independence was observed in the Shape task. Furthermore, no retrograde interference was detected in either task.

A comprehensive model of multitask perceptual learning

Examination of the session-by-session learning curves revealed task-specific anterograde interference and the absence of retrograde interference in the sequential learning of multiple tasks. Since previous studies have demonstrated that block-by-block learning curves in perceptual learning involve multiple short- and long-term component processes, including general learning, between-session gain and forgetting, and within-session rapid relearning and adaptation in perceptual learning³⁶, and that the general learning rate is significantly influenced by initial performance level, task characteristic, and individual differences³⁵, we developed a comprehensive model. This model incorporated sequence position and the determinants of general learning rate into the multi-component model³⁶ and we applied it to re-analyze learning curves in Experiments 1A and 2 at a block-by-block

level. In addition, we also added a new long-term forgetting component into this framework to model the performance during re-test (Experiment 1B).

Figure 6 depicts predictions from a simplified version of the comprehensive model for the tasks trained early or late within a sequence of seven tasks, with a re-test. We applied this model to examine the four hypotheses (Fig. 1). Specifically, if the tasks in the sequence were independent, there would be no sequence position effect (Fig. 6a). If there were facilitation between tasks, performance in the late sequence position would exhibit better initial performance and/or faster general learning rate (Fig. 6b). If there were retrograde interference, subjects trained in the early sequence position would exhibit more forgetting during re-test (Fig. 6c). If there were anterograde interference, performance in the late sequence position would exhibit both worse initial performance and/or slower general learning rate (Fig. 6d).

Modeling Experiment 1 A

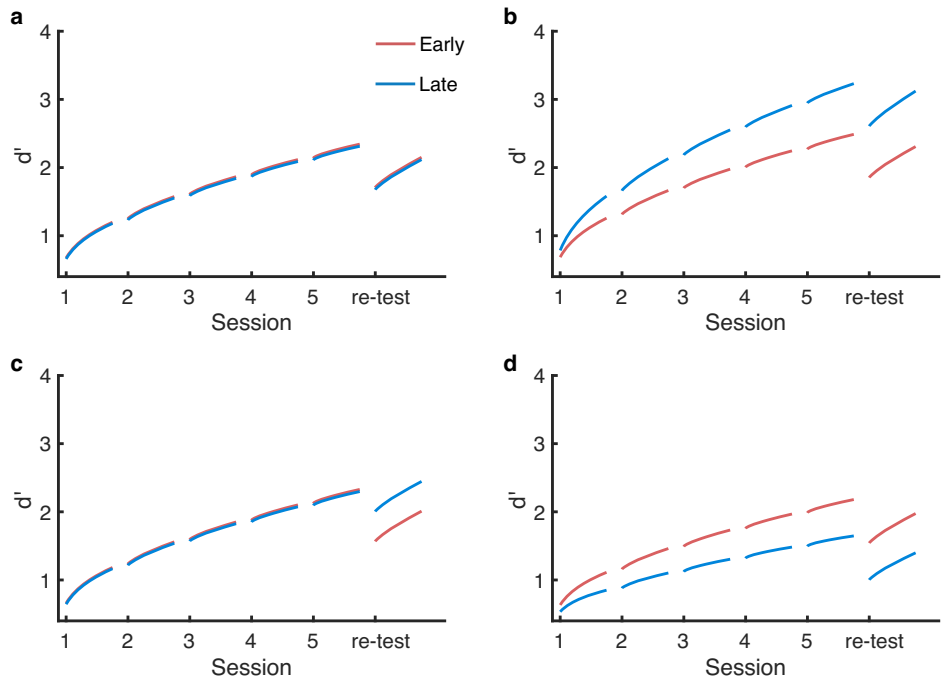
As a baseline, we first fitted the comprehensive model to the 343 block-by-block learning curves for the seven tasks from forty-nine subjects in Experiment 1A, without considering the effects of sequence position. The model consisted of a total of 96 parameters (See Methods) and accounted for a substantial proportion of variance (Mean \pm SE: $30.21\% \pm 6\%$; Supplementary Fig. 3). Subsequently, we employed the forward stepwise regression approach to assess the sequence position effect (See Methods). The best fitting model included five additional parameters: coefficients of the sequence position on the initial performance of the Auditory and N -back tasks, and on the general learning rate of the Vernier, Contrast, and Face tasks (Fig. 7b and Supplementary Fig. 4). This model demonstrated statistical equivalence to the full model (32.40% vs 32.49% , $F(9,11650) = 1.74$, $p = 0.08$, Cohen's $f^2 = 0.001$) and was significantly superior to the most reduced baseline model (32.40% vs 30.21% , $F(5,11659) = 75.52$, $p < 0.001$, Cohen's $f^2 = 0.032$).

The predictions of the best fitting model for the average block-by-block early (sequence positions 1, 2, and 3) and late (sequence positions 5, 6, and 7) learning curves are depicted in Fig. 7a, alongside the experimental data. Across the Vernier, Face, Contrast, Auditory, and N -back tasks, the average learning curves in early positions exceeded those in late positions.

The best-fitting model effectively estimated the initial performance in each task (Fig. 7c), showing high consistency with the measured initial performance obtained by averaging the performance of 49 subjects in the first block (Supplementary Fig. 5a, $r > 0.99$, $p < 0.001$). More importantly, sequence position significantly affected initial performance in the Auditory ($\lambda_1 = -0.007$) and N -back ($\lambda_1 = -0.091$) tasks, as well as the general learning rate in the Vernier ($\lambda_2 = -0.014$), Face ($\lambda_2 = -0.007$), and

Fig. 6 | Predictions of the comprehensive model.

The same task is trained either early (red curves) or late (blue curves) within a sequence of seven tasks with a re-test. **a** Independence: No sequence position effect. **b** Facilitation: Performance in the late sequence position exhibits better initial performance and/or faster general learning rate. **c** Retrograde Interference: The early sequence position exhibits more forgetting during re-test. **d** Anterograde Interference: Performance in the late sequence position demonstrates both worse initial performance and/or slower general learning rate.



Contrast ($\lambda_2 = -0.015$) tasks (Fig. 7b), aligning with the anterograde interference hypothesis (Fig. 6d).

We further estimated the coefficients of task, subject, and initial performance factors on general learning rate (Fig. 7d). While the numerical values of the coefficients for the task and initial performance factors differed from those in the original study³⁵ due to the use of raw and normalized data in the two studies, the subject-specific learning ability, remained highly consistent (Fig. 7c, Supplementary Fig. 6a, $r = 0.77$, $p < 0.001$). Additionally, the estimated short-term and long-term components derived from the comprehensive model were largely consistent with those in the original study³⁶ (Fig. 7e-g; Supplementary Fig. 6b, $r = 0.58$, $p = 0.003$).

Although the quantitative effects of sequence position on the block-by-block learning curves derived from the comprehensive model differed from those obtained through the session-by-session analysis, a consistent pattern of anterograde interference was observed in some tasks. Specifically, both analyses revealed independence in the shape search task and negative sequence position effects on the initial performance of the auditory frequency discrimination task and on the general learning rate of face view discrimination task.

For Vernier offset discrimination and contrast detection tasks, the block-by-block analysis identified sequence position effects on the general learning rate while the session-by-session analysis indicated effects on initial performance. In contrast, For the motion direction discrimination task, we found only a marginal sequence position effect in the block-by-block analysis ($p = 0.08$), but a significant effect on initial performance in session-by-session analysis (Fig. 3). Moreover, the block-by-block analysis captured large sequence position effect on initial performance in the *N*-back task (Fig. 7b), which was not evident in session-by-session analysis (Fig. 3). These findings underscore the importance of considering the temporal grain in analyzing learning curves, as it can significantly impact the conclusions drawn.

Modeling Experiment 1B

We then applied the comprehensive model to analyze data from Experiment 1B. Employing the same forward stepwise regression approach (see Methods), we utilized 21 parameters (three for each of the seven sequence positions) to evaluate the effect of sequence position on initial performance (λ_1), general learning rate (λ_2), and performance reduction between the last training session and re-test (λ_3).

The best-fitting model included coefficients of sequence position on initial performance in the Auditory, *N*-back, Vernier, Face, and Motion tasks, as well as coefficients of sequence position on the general learning rate for the Contrast, and Motion tasks (Fig. 8b). This model was statistically equivalent to the full model (35.81% vs 36.09%, $F(14,4789) = 1.48$, $p = 0.11$, Cohen's $f^2 = 0.004$) and significantly superior to the baseline model (35.81% vs 31.32%, $F(7,4803) = 48.03$, $p < 0.001$, Cohen's $f^2 = 0.07$). Most of the coefficients were negative, suggesting that the performance was better if a task was trained early in the sequence.

The predictions of the best fitting model for the average block-by-block early and late learning curves during training and re-test are depicted in Fig. 8a. Due to the notable difference between long-term forgetting (Fig. 8c) and regular between-session effects (Fig. 8f), our model included separate parameters for within-session relearning for each task (Fig. 8h) and within-session adaptation in the contrast detection task during the re-test (Fig. 8i). The model accurately captured the learning curves for both the five initial training sessions and the re-test months later, with distinctive variations of the curves, particularly in the abrupt changes in the shape search and *N*-back tasks. Similar to the modeling results in Experiment 1A, the estimated initial performance for the seven tasks in Experiment 1B (Fig. 8d) closely aligned with the measured data (Supplementary Fig. 5b, $r = 0.97$, $p = 0.002$).

In line with the observation in Experiment 1A, where sequence position influenced Vernier, Face, Contrast, Auditory, and *N*-back tasks, Experiment 1B revealed a similar pattern. Here, early sequence positions consistently exhibited higher average learning curves compared to their late sequence counterparts. The best-fitting model revealed that sequence position impeded the initial performance in the Auditory ($\lambda_1 = -0.019$), *N*-back ($\lambda_1 = -0.240$), Vernier ($\lambda_1 = -0.009$), and Face ($\lambda_1 = -0.015$) tasks, as well as the general learning rate in the Contrast ($\lambda_2 = -0.009$) task. Notably, the best-fitting model also indicates that sequence position exerted a negative influence on the general learning rate ($\lambda_2 = -0.048$) but a positive influence on the initial performance ($\lambda_1 = 0.085$) of the Motion task.

Importantly, this superior performance of early learners extended beyond the five training sessions and continued into the re-test session. Additionally, we observed ubiquitous long-term forgetting across all tasks (Fig. 8c), unaffected by sequence position, substantiated by the best-fitting model which was devoid of sequence position-related coefficients for long-

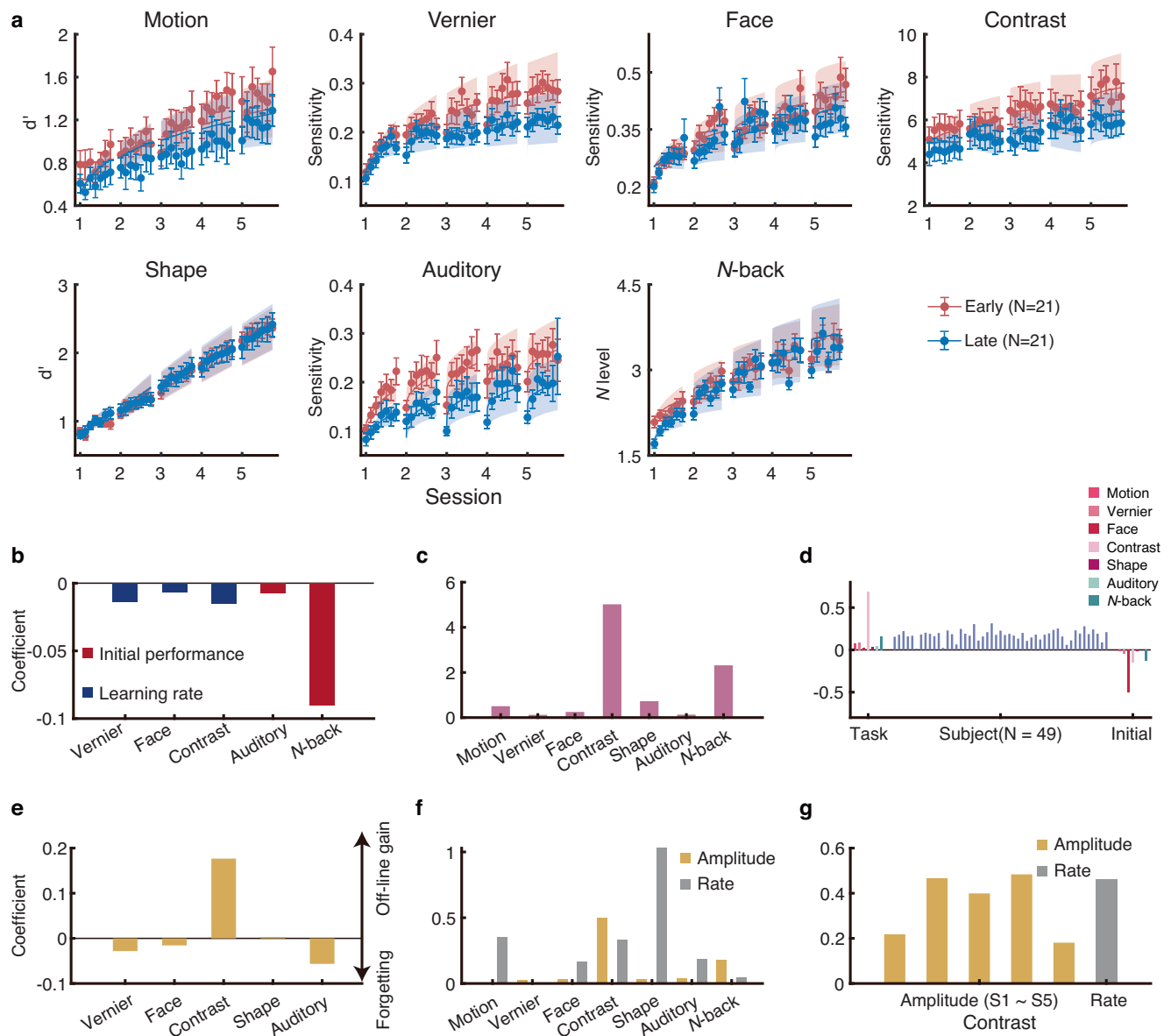


Fig. 7 | Best fitting comprehensive model (Experiment 1A, $N = 49$). **a** Predictions of the best fitting model for the average block-by-block early (red: sequence positions 1, 2, and 3) and late (blue: sequence positions 5, 6, and 7) learning curves for the seven tasks, along with the experimental data. Estimated coefficients of sequence position (**b**) and initial performance (**c**). Estimated coefficients of factors that influence general learning rate (**d**): 7 task-specific parameters (left), 49 subject-specific parameters (middle), and 7 task-specific coefficients reflecting the influence of initial

performance. Each task was denoted by one color. **e** Estimated parameters for between-session effects in five tasks. **f** Estimated parameters for within-session relearning in all seven tasks, with yellow bars indicating the amplitude of relearning and gray bars indicating the rate of relearning. **g** Estimated parameters for within-session adaptation. The five yellow bars represent the amplitude of adaptation in each session, while the gray bar indicates the rate of adaptation.

term forgetting (Fig. 8b). These findings corroborated the presence of anterograde interference rather than retrograde interference in learning these seven tasks (Fig. 6). Furthermore, the coefficients representing the effect of task, subject, and initial performance on general learning rate show strong consistency between the 17 subjects and the broader sample of 49 subjects (Supplementary Fig. 7a–c, task: $r = 0.94$, $p = 0.002$; subject: $r = 0.96$, $p < 0.001$; initial performance: $r = 0.67$, $p = 0.10$). The coefficients of the short-term components also aligned with those in Experiment 1A (Fig. 8f, g, i; Supplementary Fig. 7d, $r = 0.73$, $p < 0.001$), validating the stability of the comprehensive model.

In the session-wise analysis, sequence position influenced the initial performance of Vernier, Auditory, and N-back tasks, which were replicated in the comprehensive model analysis (refer to Fig. 4c and Fig. 8b). In

addition, the comprehensive model uncovered negative effects of sequence position on the general learning rate of the Contrast task and the initial performance of the Face task, which were not evident in session-wise analysis. However, the model analysis from Experiment 1A confirmed the negative influence of sequence position on these two tasks, underscoring the importance of a finer analysis scale with an ample participant pool.

Modeling Experiment 2

Finally, we applied the comprehensive model to analyze the learning curves in Experiment 2. Since each subject in this experiment was only engaged in two tasks, extracting the subject-specific general learning ability was not feasible. Consequently, we replaced the subject factor with a group factor (Fig. 9d) to capture the differences between the two groups. Results revealed

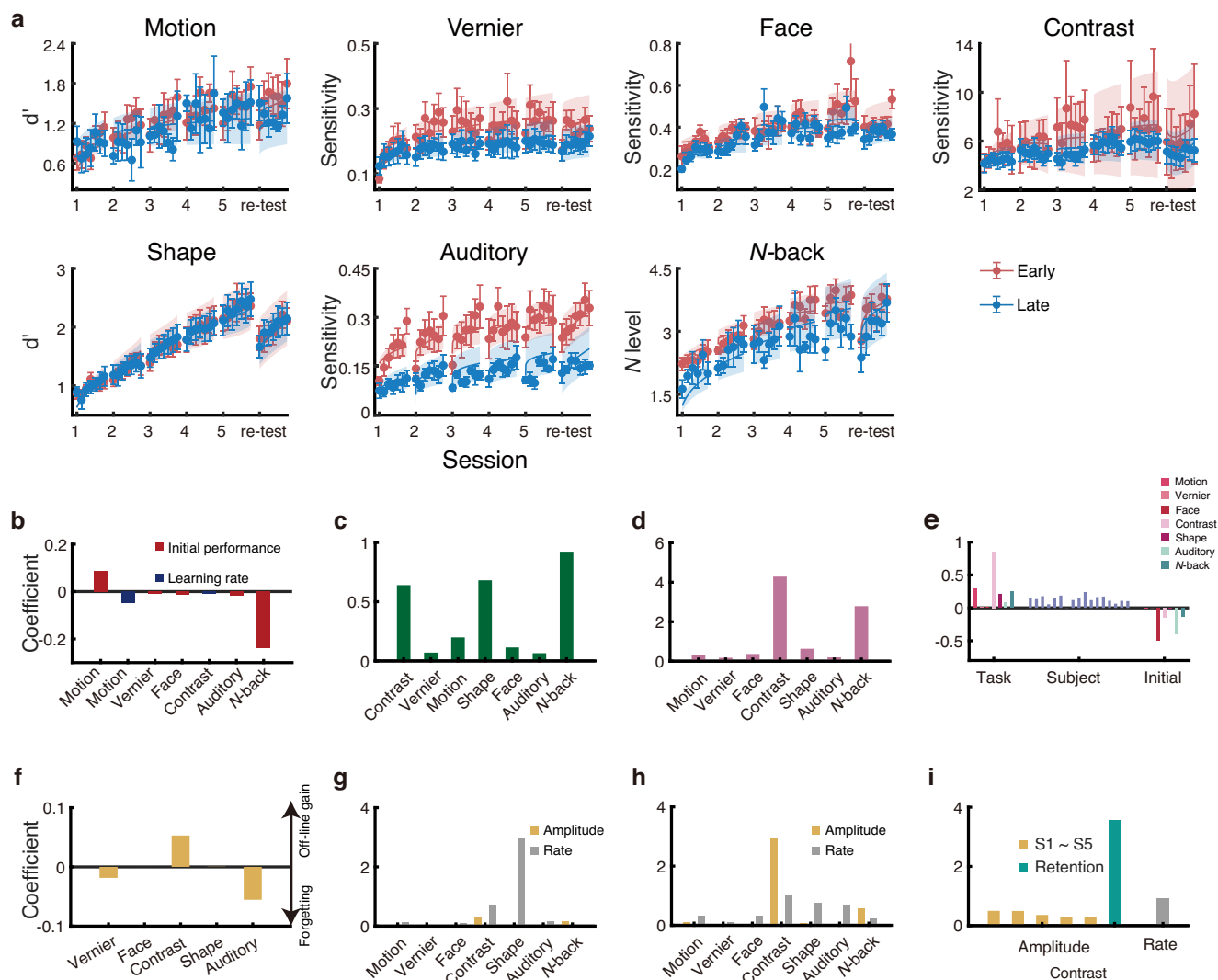


Fig. 8 | Retrograde interference analysis with the comprehensive model (Experiment 1B, $N = 17$). **a** Predictions of the best fitting model for the average block-by-block early (red: sequence positions 1, 2, and 3) and late (blue: sequence positions 5, 6, and 7) learning curves for the seven tasks, along with experimental data. Estimated coefficients of sequence position (**b**), long-term forgetting (**c**), initial performance (**d**). Estimated coefficients of factors influencing general learning rate (**e**): 7 task-specific parameters (left), 17 subject-specific parameters (middle), and seven task-specific coefficients reflecting the influence of initial performance.

f Estimated parameters for between-session effects in five tasks. **g** Estimated parameters for within-session relearning during 5 training sessions, with yellow and gray bars indicating the amplitude and rate of relearning, respectively. **h** Estimated parameters for within-session relearning during re-test, with yellow and gray bars indicating the amplitude and the rate of relearning, respectively. **i** Estimated parameters for within-session adaptation in the contrast detection task. The yellow bars and the green bar represent the amplitude of adaptation in the training sessions and the retention session, respectively, while the gray bar indicates the rate of adaptation.

that including a sequence position coefficient for initial performance on the Vernier task (Fig. 9) led to a statistically equivalent fit as the full model (30.07% vs 30.09%, $F(3,1101) = 0.11$, $p = 0.95$, Cohen's $f^2 = 0.0003$), and a significantly better fit than the most reduced model (30.07% vs 29.76%, $F(1,1104) = 4.87$, $p = 0.03$, Cohen's $f^2 = 0.0044$).

The predicted block-by-block learning curves are depicted in Fig. 9. As illustrated in Fig. 9a, the group that learned the Vernier task first exhibited better performance, supported by a negative coefficient of sequence position on its initial performance (Fig. 9b, $\lambda_1 = -0.014$). In contrast, the group that learned the shape task at the second sequence position (Group 1) showed better performance than the other group (Group 2). The pattern may suggest some form of facilitation (Fig. 9b). However, upon considering the group factor, we found that the patterns of learning curves for the shape task were reflective of different general learning abilities of the two groups (Fig. 9d: Group 1 vs Group 2: 0.033 vs 0.008) rather than sequence position effects (Fig. 9b). Additionally, despite Experiment 1A and Experiment 2 involving different samples of subjects, the task-related components estimated by our

comprehensive model showed high consistency ($r = 0.80$, $p = 0.002$; Fig. 9c–f and Supplementary Fig. 8).

Discussion

We aimed to investigate interactions across multiple perceptual tasks learned through sequential training, formulating four hypotheses: independence, facilitation, retrograde interference, and anterograde interference. These hypotheses were tested using two carefully structured experiments to measure sequence position effects.

In Experiment 1, 49 subjects learned seven tasks sequentially in seven distinct training sequences (Experiment 1A), with a subset of them re-tested months later (Experiment 1B) to evaluate retrograde interference. Experiment 2 involved 14 new subjects learning two of the seven tasks sequentially, followed by a re-test on the first trained task. Using session-by-session analysis, we found that later sequence position had a negative influence on the initial performance of motion sequence direction discrimination, Vernier offset discrimination, contrast detection, and auditory frequency discrimination

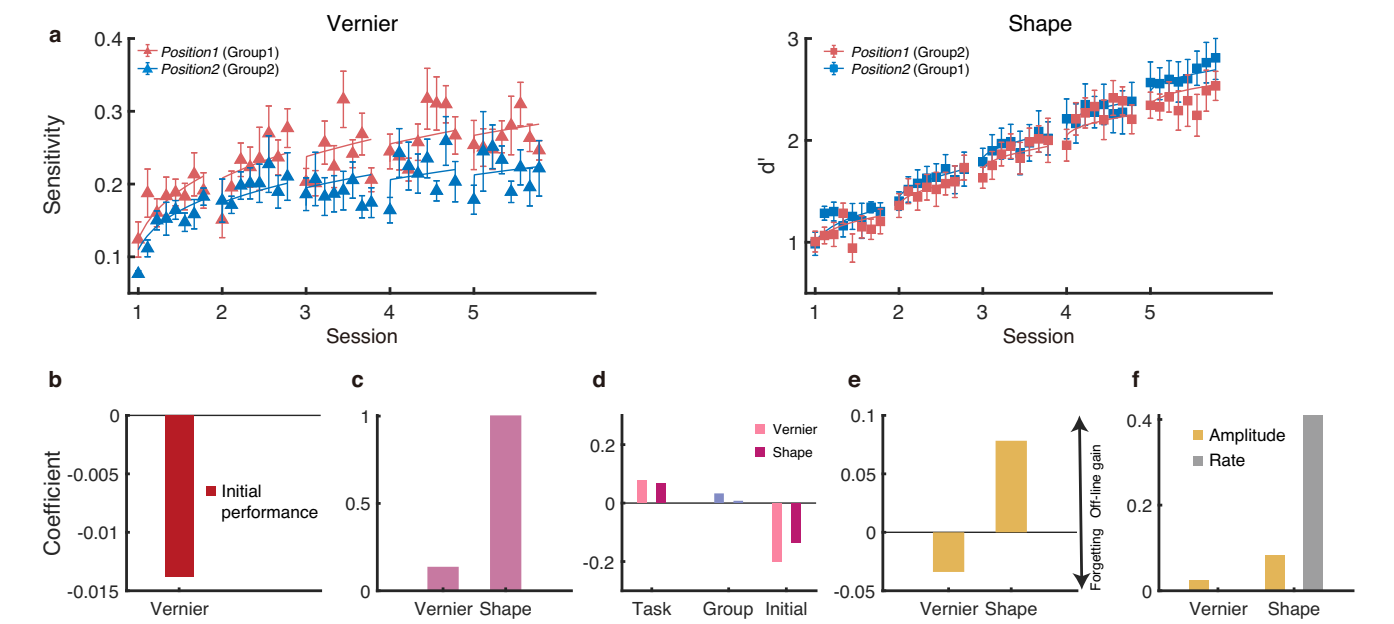


Fig. 9 | Task-specific anterograde interference analysis with the comprehensive model (Experiment 2, N = 14). **a** Predictions of the best fitting model for the average block-by-block learning curves in the first (red) and second sequence position (blue) for the two tasks, along with experimental data. Estimated coefficients of sequence position (**b**), initial performance (**c**). Estimated coefficients of factors influencing general learning rate (**d**): two task-specific parameters (left), two group-specific parameters (middle), and two task-specific coefficients reflecting the influence of initial performance (right). **e** Estimated parameters for between-session effects. **f** Estimated parameters for within-session relearning, with yellow and gray bars indicating the amplitude and the rate of relearning, respectively.

Table 1 | Sequence position effects at both session-wise LMMs and block-wise model analysis

Task	Experiment 1 A (N = 49)		Experiment 1 B (N = 17)		Experiment 2 (N = 14)	
	Session	Block	Session	Block	Session	Block
Motion	Initial (↓)	\	\	Initial (↓) & Rate (↑)		
Vernier	Initial (↓)	Rate (↓)	Initial (↓)	Initial (↓)	Initial (↓)	Initial (↓)
Face	Rate (↓)	Rate (↓)	Initial (↓)	Initial (↓)		
Contrast	Initial (↓)	Rate (↓)	\	Rate (↓)		
Shape	\	\	\	\	\	\
Auditory	Initial (↓)	Initial (↓)	Initial (↓)	Initial (↓)		
N-back	\	Initial (↓)	\	Initial (↓)		

“Initial” represents sequence position influence on the initial performance while “Rate” represents sequence position influence on the general learning rate for that task. “↓” represents negative influence and “↑” color represents positive influence. “\” represents sequence position exerts no influence on that task.

tasks, while it also slowed the learning in the face view discrimination task (Experiment 1A). The re-test in Experiment 1B showed sustained negative effects of later sequence position on the performance at the last training session in Vernier offset discrimination, face view discrimination, and auditory frequency discrimination tasks, but no effects on forgetting. Experiment 2 confirmed that sequence position impeded the initial performance of Vernier offset discrimination task. Intriguingly, the shape search task remained unaffected by later sequence position in both experiments.

Subsequently, we developed a comprehensive model to jointly model sequence position effects, along with many factors that affect the general learning rate³⁵ and multiple short-term and long-term component processes³⁶. This model was then applied to perform block-by-block analysis of the data from both experiments. In Experiment 1A, we found that sequence position had a negative influence on the initial performance of Auditory and N-back tasks, as well as the general learning rate of Vernier, Contrast, and Face tasks. In Experiment 1B, these negative effects of later sequence position on the initial performance of Auditory and N-back tasks, and on the general learning rate of the Contrast task, were replicated.

Additionally, sequence position negatively impacted the initial performance of Vernier and Face tasks, and impeded the general learning rate but facilitated the initial performance of Motion task. In Experiment 2, sequence position consistently imposed a detrimental role on the initial performance of the Vernier task. Interestingly, the impact of sequence position on the Shape task remained consistent across subject sizes and various analyses, whether analyzed in session-wise or block-wise analysis.

Although the two types of analysis yielded different quantitative results, they were consistent in terms of the major effects (Table 1), particularly regarding the negative effect of later sequence position on tasks such as Vernier, Face, Contrast, Auditory, and N-back tasks. The block-wise approach provided a more sensitive measure of sequence position effects on the general learning rate, whereas the session-wise analysis might attribute this effect to initial performance due to averaging over rapid learning processes, echoing findings from ref. 17.

These findings indicate that training on a specific task can both enhance performance on that task and cause anterograde interference in subsequent tasks, though some tasks may remain unaffected. Previous studies have primarily reported short-term anterograde interference when

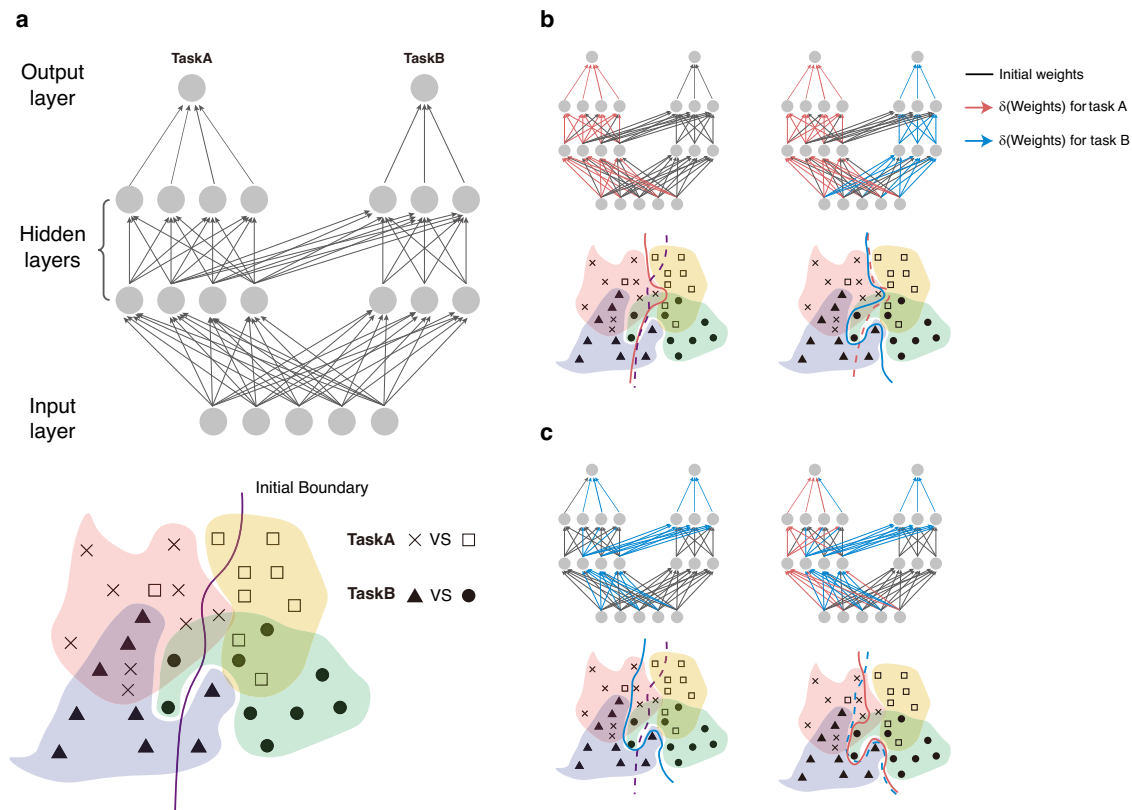


Fig. 10 | A schematic representation of a neural network that can yield asymmetric interactions in learning two tasks. a The initial state of a neural network before learning task A and B (top), along with the model's initial decision boundary on the four features involved in the two tasks (bottom). **b** The state of the network

after learning task A first and task B second (top) and the corresponding decision boundaries (bottom). **c** The state of the network after learning task B first and task A second (top) and the corresponding decision boundaries (bottom). Red represents change for task A and blue for task B.

two tasks are training consecutively within short intervals, emphasizing the role of the post-training period in consolidating existing memory traces and forming new memories^{19,37,38}. Magnetic resonance spectroscopy (MRS) studies have shown that rapid neurochemical processes, particularly the excitatory-to-inhibitory (E/I) ratio of glutamate to GABA, regulate brain state transitions between plasticity and stability^{19,39}. An increased E/I ratio promotes a plastic brain state conducive to memory formation, while a rapid decrease in the E/I ratio fosters a stable state for memory consolidation, which can hinder new task learning and lead to anterograde interference. In contrast, our study spaced training sessions by at least 6 h—often across separate days—a duration sufficient for memory consolidation²⁰. Despite this interval, which we expected would minimize short-term interference, we consistently observed robust interference, suggesting a competitive process involving long-term memory traces. In both experiments, the Vernier task exhibited anterograde interference, while the Shape task remained independent. Although asymmetric interference is rarely reported in interference studies, it aligns with facilitation, the second interaction we have proposed (Fig. 1b). For instance, McGovern et al. found that training on an orientation discrimination task improved performance on a curvature task but less on a global form task, while training on the global form task enhanced curvature task performance but not orientation task performance³². Despite their opposing behavior outcomes, facilitation and interference share underlying processes⁴⁰, supporting the notion that perceptual learning across tasks involves distinct plastic sites with compatible or incompatible overlapping weight structures⁴¹.

The asymmetric interference pattern across tasks of varying complexity suggests that perceptual learning engages distinct neural sites with task-specific weights. Low-level tasks, such as Vernier offset discrimination, primarily activate lower-level regions (e.g., V1 and V4⁴²), while visual shape search tasks involve both early (retinotopic cortex) and higher-level areas

(lateral occipital cortex)⁴³. This may create incompatible weight structures in the primary visual cortex, though shape search tasks, relying on flexible higher-level regions, may reduce interference by adjusting connections between these areas. Future research integrating deep neural network modeling with fMRI could identify overlapping brain regions, deepening insights into the neural mechanisms of anterograde interference.

We devised a neural network architecture to elucidate the asymmetrical anterograde interference (Fig. 10). The network comprises an input layer, hidden layers, and an output layer with output nodes for task A and task B (Fig. 10a). The two tasks share some nodes in the input and hidden layers on the left side of the network, with additional hidden units on the right side specific to task B. Prior to learning, the decision boundary between the four types of features used in the two tasks is able to support some level of performance in both tasks. If the model learns task A first, all the relevant connection weights for task A are adjusted to enhance performance in this task (Fig. 10b). Consequently, the decision boundary is optimized for distinguishing the two features used in task A. To ensure that subsequent learning of task B does not degrade performance in task A, the learned task A-specific weights are protected from further modification. Instead, the weights involving only non-overlapping nodes on the left side of the network and the additional hidden units on the right side of the network are modified to enable learning of task B. This allows task B to be learned without being affected by sequence position, reflecting our observations on the shape search task.

In contrast, if the model learns task B first, all the relevant weights for task B are adjusted to enhance performance in that task (Fig. 10c). Thus, the decision boundary is optimized for learning task B. Similar to the previous scenario, the learned task B-specific weights are protected from further modification to maintain performance in task B when learning task A. However, the weights involving only non-overlapping nodes on the left side are allowed to change, which prevents the network from developing the

optimal decision boundary for task A, resulting in anterograde interference similar to those observed in the experiments.

The current study suggests that models of perceptual learning might incorporate shared and independent hidden layers that preserve weight structures from learned tasks. The Integration Reweighting Theory and related reweighting models propose that learning involves optimizing connection weights between stable stimulus representation and task decisions^{44,45}. While these theories and models have primarily been applied to investigate perceptual learning and transfer in single tasks, our investigation of sequential learning across multiple tasks may involve shared and independent hidden layers as well as preservation of the weight structure from learned tasks. Such kind of network is crucial for understanding task interactions, particularly in the case of asymmetric effects, as well as for developing accurate models capable of acquiring multiple perceptual expertise.

Furthermore, these findings hold significant implications for developing artificial networks capable of learning multiple tasks. While neural networks excel in single-task performance⁴⁶, artificial general intelligence struggles with catastrophic forgetting when learning multiple tasks sequentially²¹. Our results revealed that while human memory experiences natural decay over time⁷, it does not suffer retrograde interference due to being trained in a later sequence position. This finding is also consistent with existing literature^{20,38}, which demonstrates that a time interval of several hours between two training sessions is sufficient for memory to consolidate. In our study, the training sessions were typically spaced a day apart or, at minimum, separated by 6 h. Inspired by human memory processes, several studies have implemented strategies such as strengthening old memories through techniques like replay^{47,48} and off-line consolidation²², or dynamically adjusting plasticity^{21,49} to avoid forgetting. Our results supported the potential meta-plasticity strategy for balancing plasticity and stability: reducing plasticity in previously engaged neural regions as a protective mechanism for preserving established learning. This strategy could be more energy-efficient and task-autonomous⁴⁹ particularly when learning a considerable number of tasks sequentially.

Our current study has provided compelling evidence for the presence of anterograde interference when learning multiple tasks sequentially, even when considering multiple factors affecting general learning process and the detailed temporal dynamics of the learning process. Asymmetric interference suggests that perceptual learning engages distinct plastic sites with compatible or incompatible overlapping weight structures⁴¹, necessitating meta-plasticity mechanism to balance between plasticity and stability. Tasks reliant on low-level sensory features were more prone to interference, possibly due to greater adjustments in overlapping neural pathways. Alternatively, overtraining to an expert level may automate processing, causing synaptic pruning^{50,51}, which could reduce overlap and interference. These studies highlight that synaptic plasticity may be energy-efficient and task-autonomous during multitask learning⁴⁹.

While extensive studies have focused on mechanisms underlying learning a single task, extending this inquiry to the dynamic processes involved in multitask learning could yield significant insights into the plasticity and stability balance of the human brain, especially considering the complex environment and multifaceted tasks we encounter in everyday life⁶. The observed task-specific anterograde interference in our study could provide valuable inspiration for designing training protocols aimed at enhancing training efficiency. For instance, placing low-level perceptual tasks early in the training sequence, particularly for clinical and expertise training requiring the acquisition of multiple proficient skills, could optimize learning outcomes. By integrating our behavioral results and ongoing investigations into the neural networks of the human brain⁵², we can deepen our understanding of the strategies employed by the biological brain to navigate the complex and dynamic real world.

Methods

Subjects

A total of 63 naïve subjects recruited from Chinese universities took part in the study. Experiment 1A involved 49 young adults (22 males, age range 20–30, mean age 23 years old), with 18 of them returning for a follow-up

retention test (Experiment 1B) after 3 to 9 months (Mean \pm SE: 6.67 \pm 1.78 months). One subject from Experiment 1B was omitted from the analysis represented in the main text due to the estimated threshold in the auditory discrimination task being more than five standard deviations lower than the average performance of all the other subjects in the last training session of the task. Analysis including this subject in the Supplementary materials yielded consistent results with the main text. The remaining 14 undergraduate or graduate students (7 males, age range 18–28, mean age 20 years old) took part in Experiment 2. All subjects provided informed consent prior to the studies and were paid for participation. They all had normal or corrected-to-normal vision and wore their corrective glasses, if necessary. None reported any psychiatric or neurological disorder. The study was approved by the Ethics Committee of Institute of Psychology, Chinese Academy of Sciences (Project ID Number: H20058). The facial images used in this study were generated using FaceGen software and are entirely synthetic. As these images do not depict any real individuals and do not contain any identifying information, obtaining written informed consent was not applicable.

Apparatus

Two SONY G220 color monitors, with a resolution of 1600 \times 1200 pixels and a refresh rate of 85 Hz and a DELL E1912Hc LCD monitor, were used in the study. To enhance gray-level resolution, we employed a specialized circuit that combined two 8-bit output channels of the graphics card, hence enabling a 14-bit gray-level resolution⁵³. The experiment was programmed using MATLAB (Mathworks, Natick, Massachusetts) with PsychToolbox extensions^{54,55}.

Experimental design

In Experiment 1A, all subjects underwent sequential training of seven tasks, including motion direction discrimination, Vernier offset discrimination, face view discrimination, contrast detection, visual shape search, auditory frequency discrimination, and audiovisual *N*-back working memory tasks. Each task was trained in five consecutive sessions, each lasting ~40 to 60 min (Fig. 2b). Most sessions were conducted on separate days, although in some cases, two consecutive sessions occurred on the same day with a minimum interval of six hours between them²⁰. For Vernier offset discrimination, face view discrimination, contrast detection, and auditory frequency discrimination tasks, we used a single staircase method²⁶. The initial parameter was identical for all subjects in the first block of the first session, set to be easy for quick adaptation. Subsequent sessions started at 1.2 times the previous day's average threshold, and within sessions (except the first block), the starting value matched the previous block's final trial. For global motion discrimination and visual shape search tasks, we used the constant stimulus method. For the *N*-back task, we applied a modified paradigm⁵⁶ over 5 days, with 2 daily mini-sessions of 15 mini-blocks each. The *N* level began at 1 for the first 3 mini-sessions and 2 for the remaining seven. The *N*-back task included 30 blocks, each with 20 + *N* trials; the other six tasks had 7 blocks per session, with 96 or 100 trials per block.

In Experiment 1B, subjects completed seven re-test sessions, adhering to their original trained sequence, typically on separate days, with a minimum 6-h interval if on the same day. For Vernier offset discrimination, face view discrimination, contrast detection, and auditory frequency discrimination tasks, the initial parameter of the first trial in the first block during the re-test was set to the same value as at the start of training. For the remaining six blocks, the starting value of each block's first trial was the previous block's final trial value, using a 3-down/1-up staircase⁵⁷ to adjust stimulus levels within blocks. This consistent initial parameter across subjects facilitated quick re-familiarization with easy trials. For the *N*-back task, subjects completed two mini-sessions during the re-test: the first started at *N* = 1, the second at *N* = 2, with *N* adjusted per mini-block based on performance. For global motion discrimination and visual shape search tasks, we used the same constant stimuli method as in training.

In Experiment 2, 14 subjects were trained on the visual shape search and the Vernier offset discrimination tasks sequentially, each across five

daily sessions. Half began with the Vernier task, while the other seven started with the Shape task. Upon completing the second task, subjects underwent re-test on the first trained task on the following day. For both tasks, each session involved 8 blocks of 96 trials.

Tasks

In the global motion discrimination task, subjects performed a two-interval-forced-choice (2IFC) global motion direction discrimination task (Fig. 2a). The task involved observing two sets of 400 moving dots ($0.18^\circ \times 0.18^\circ$ each) with the same speed ($10^\circ/\text{s}$), same circular aperture (8° in diameter), and either the same (both at 0°) or different (0° vs 2.5° or -2.5°) moving directions. Subjects reported whether the moving directions of the two stimuli were identical or different, with the correct response accompanied by an auditory beep. The probability of the dots moving in either identical or varying directions was evenly balanced. Performance was assessed by monitoring the percentage of correct responses.

In the Vernier offset discrimination task, subjects performed a two-alternative-forced choice (2AFC) Vernier offset discrimination task (Fig. 2a). The Vernier stimulus, consisting of two Gabor stimuli (contrast = 0.45, spatial frequency = 3 cycle per degree, and $\sigma = 0.29^\circ$), was presented at 5° retinal eccentricity in the upper left visual quadrant for 200 ms. A slight position jitter (within 0.25°) was added to the stimulus position in each trial. To ensure fixation, nine black letters ($1.56^\circ \times 1.56^\circ$) were displayed sequentially in the fovea in the stimulus interval. Subjects reported both the foveal letter (H or N) and the offset direction of the Vernier stimulus: whether the lower Gabor was to the left or right of the upper Gabor, with an auditory beep followed each correct response. The offset of the Gabor was controlled by a 3-down/1-up staircase that converges to 79.4% correct⁵⁷, and the step size of the offset change was set to 10%. The threshold for each block was computed by calculating the arithmetic mean of the remaining even number of reversals after excluding the first four or five reversals within that block. The initial offset was 12.5 arcmin for all subjects.

In the face view discrimination task, a three-dimensional (3D) face model without hair was obtained from FaceGen Modeler 3.1 (<http://www.facegen.com/>). This 3D face model was then rotated along various in-depth angles to the monitor plane, with the front view (0°) as the initial position. This rotation generated a range of face-view stimuli⁵⁸. Subjects performed a 2IFC face view discrimination task (Fig. 2a). One 100-ms stimulus interval contained a face at 30° and the other stimulus interval contained a face at $30^\circ \pm \theta^\circ$, with the order randomized and position jittered within a $1.43^\circ \times 1.43^\circ$ area across trials. Subjects reported whether the second face was titled left or right relative to the first face, with auditory feedback following each correct response. The initial difference was 8° and a 3-down/1-up staircase controlled the difference between the two face views⁵⁷.

In the contrast detection task, subjects performed a 2IFC contrast detection task (Fig. 2a). The task involved detecting a target vertical sinusoidal grating (with a spatial frequency of 24 cycle/degree and a random phase), which occupied a $2^\circ \times 2^\circ$ visual angle and was windowed with a Gaussian ramp ($\sigma = 0.25^\circ$). Subjects were required to judge which interval contained the grating, with correct response accompanied by an auditory beep. Stimulus contrast was controlled by a 3-down/1-up staircase⁵⁷. The initial contrast level was set at 0.60.

In the visual shape search task, subjects performed a Yes-or-No visual shape search task. The stimuli consisted of a central fixation dot and 24 triangles of four possible orientations (left, right, up, and down) and were evenly spaced on a 5×5 grid ($3.42^\circ \times 3.42^\circ$; Fig. 2a)⁵⁹. After the 800 ms stimulus presentation, subjects reported if they detected a triangle of downward orientation (target) among the 24 triangles, with auditory feedback following each response. The probability of target presence was 0.75 and the target was evenly presented to the 24 possible locations in the target-presence trials. Percent correct was monitored throughout training.

In the auditory frequency discrimination task, subjects performed a 2IFC auditory frequency discrimination task (Fig. 2a) binaurally with Sennheiser HD600 headphones in a quiet room. One stimulus interval contained a 100-ms tone with a fixed frequency of 1000 Hz, and the other

interval, following an inter-stimulus interval of 500 ms, contained a 100-ms tone with a frequency of $1000 + \Delta$ Hz. The sequence of the two tones was randomized. Subjects reported which interval contained the higher tone, with 100-ms visual feedback after each response indicating the accuracy of their response. Tone bursts, including 10-ms rise-falls, were generated with a 16-bit digital-to-analog converter (sample rate = 44.1 kHz) and modulated by a raised cosine function. A 3-down/1-up staircase controlled the frequency difference between the two tones⁵⁷. The initial frequency difference was 30 Hz for all subjects.

In the audiovisual *N*-back working memory task, a sequential audiovisual *N*-back paradigm was adopted⁶⁰. The auditory materials were delivered via Sennheiser HD600 headphones. In each trial, subjects simultaneously viewed a visuospatial stimulus (a $2^\circ \times 2^\circ$ blue square) displayed for 500 ms at one of eight screen locations, and listened to an English consonant that was selected from the set [C, H, K, L, Q, R, S, T] (Fig. 2a). The inter-trial interval was 2500 ms, during which subjects judged if the current visual/auditory stimulus matched the one from *N* trials prior. No response was required for non-target trials. A mini-block included $20 + N$ trials and 6 targets per modality, occurring randomly in the two stimulus streams. The *N* level was disclosed before the start of each mini-block. If a subject erred less than three times in both modalities, *N* was increased by 1 in the next mini-block; if five or more errors were registered, *N* was decreased by 1. Otherwise, *N* remained unchanged. The *N*-back training spanned 5 days, consisting of 30 mini-blocks each day. In the first 3 mini-sessions, the level of *N* started from 1; all other mini-sessions started with *N* = 2.

Learning performance transformation

For contrast detection, Vernier acuity, face view discrimination, and auditory frequency discrimination tasks, block-by-block and session-by-session thresholds were estimated using a maximum likelihood estimation method^{36,61}. The estimated thresholds were then transformed into perceptual sensitivities by taking their reciprocals. For motion direction discrimination and visual search tasks, we assessed performance in terms of *d'* derived from percent correct responses. In the *N*-back working memory task, we evaluated performance for each block by calculating the average *N*-level across five mini-blocks. The performance for each session was determined by averaging *N*-level from every three blocks.

Session-wise analysis with LMMs

In order to take individual's performance into consideration when analyzing the impact of sequence position, we fitted four fixed effect models using MATLAB function *fitlme* and conducted model comparison with function *compare*. The *fitlme* function uses M-estimation to formulate the model equations and solves them through iterative reweighted least squares (IRLS). The *compare* function enables the comparison of two nested linear mixed-effects models with a likelihood ratio test. We used a simulated likelihood rate test with 1000 replications to estimate the 95% confidence interval.

The learning curves have been more accurately depicted by power law or exponential forms, rather than a linear model⁶². Hence, we logarithmically transformed performances (*sensitivity*, *d'*, *N-level*) and training sessions to facilitate the fitting and comparing of learning curves using LMMs. For each task, we fitted four models:

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} + \log_{10}(j) \times \text{slope} \quad (1)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} \times \beta_1 \times \text{position}(s) + \log_{10}(t) \times \text{slope} \quad (2)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} + \log_{10}(t) \times \text{slope} \times \beta_2 \times \text{position}(s) \quad (3)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} \times \beta_1 \times \text{position}(s) + \log_{10}(t) \times \text{slope} \times \beta_2 \times \text{position}(s) \quad (4)$$

$\hat{y}_{s,j}$ represents the predicted performance of the *s*th subject (*s* = 1, 2, ..., 49) during the *j*th session (*j* = 1, 2, 3, 4, 5). *intercept* signifies initial

performance, while *slope* represents the learning rate. Coefficients β_1 and β_2 denote effects of sequence *position* on the *intercept* and *slope*, respectively. Although we only modeled the average performance of the participants in each group, the raw performance data from all subjects were utilized.

In M1, the most reduced model, comprised only two parameters (*intercept* and *slope*). In M2, we introduced an additional parameter (β_1) to represent the effect of sequence *position* on *intercept*. For M3, we also included three parameters. Here, β_2 represents the effect of sequence *position* on *slope*. M4, the full model, consisted of four parameters, in which both initial performance and learning rate are functions of sequence *position*.

For the retention experiment, we fitted four similar models to the performance of 17 subjects in the 5th training session and the retention session: N1, there is no influence of sequence *position* on performances; N2, the effect of sequence *position* is a function of initial performance, which is the performance in the 5th session; N3, the effect of sequence *position* is a function of the magnitude of forgetting, which is the performance difference between the 5th and re-test session; N4, the effect of sequence *position* is influenced by both the performance in the 5th session and the magnitude of forgetting.

Here, the variable *intercept* and *slope* represent the initial performance (estimated from the 5th session) and the magnitude of forgetting (estimated from the difference between the 5th session and the re-test session). The two coefficients γ_1 and γ_2 represent the influence of sequence *position* on the *intercept* and *slope*.

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} + \log_{10}(j) \times \text{slope} \quad (5)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} \times \gamma_1 \times \text{position}(s) + \log_{10}(t) \times \text{slope} \quad (6)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} + \log_{10}(t) \times \text{slope} \times \gamma_2 \times \text{position}(s) \quad (7)$$

$$\log_{10}(\hat{y}_{s,j}) = \text{intercept} \times \gamma_1 \times \text{position}(s) + \log_{10}(t) \times \text{slope} \times \gamma_2 \times \text{position}(s) \quad (8)$$

Model comparison

An *F* test was employed to statistically compare the goodness of fit between any two nested models:

$$F(df_1, df_2) = \frac{(R_{full}^2 - R_{reduced}^2) / df_1}{(1 - R_{full}^2) / df_2} \quad (9)$$

where $df_1 = k_{full} - k_{reduced}$, $df_2 = n - k_{full}$, k_{full} and $k_{reduced}$ are the numbers of parameters of the full and reduced models, respectively. n is the number of data points, and R_{full}^2 and $R_{reduced}^2$ represent the goodness of fit of the full and reduced models, respectively. If the reduced model is found to be statistically equivalent to the full model, the omitted factors in the reduced model are deemed as redundant. Otherwise, the superiority of the full model would suggest that those factors are necessary in explaining the data.

Block-wise analysis with comprehensive model

To model the impact of sequence *position* on both initial performance and learning rate, we integrated the multicomponent model³⁶ and multivariate model developed in ref.³⁵ to fit raw performances in each block. Specifically, we model the general learning component in the multicomponent model as functions of subject, task, and task-specific initial performance. Next, we assumed that all subjects shared the same short-term components derived from the best-fitting model based on the averaged learning curves for each

task, based on previous work³⁶. Thus, the model is written as:

$$\begin{aligned} \hat{T}(t, s, j, k) &= \alpha(t) * (\text{block}(j-1) + k)^{\tau(t,s)} + s1(t, j) \\ &+ s2(t, j)(k-1)^{\tau_{s2}(t,j)} - s3(t, j)(k-1)^{\tau_{s3}(t,j)} \\ \tau(t, s) &= \text{task}(t) + \text{subject}(s) + \text{coeff}_{ini}(t) * \alpha(t) \\ s1(t, j) &= 0, g, g, g, g (\text{for } j = 1 \sim 5) \\ s2(t, j) &= 0, r, r, r, r (\text{for } j = 1 \sim 5) \\ s3(t, j) &= a1, a2, a3, a4, a5 (\text{for } j = 1 \sim 5) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma (\text{for } j = 1 \sim 5) \\ \tau_{s3}(t, j) &= \beta, \beta, \beta, \beta, \beta (\text{for } j = 1 \sim 5) \end{aligned} \quad (10)$$

where t represents the t th task ($t = 1, 2, \dots, 7$), s represents the s th subject ($s = 1, 2, \dots, 49$), j represents the j th session ($j = 1, 2, \dots, 5$), k represents the k th block in the current session ($k = 1, 2, \dots, 6/7$), and *block* represents the number of blocks in each session (*block* = 6 for *N*-back task, and = 7 for the other 6 tasks).

$\hat{T}(t, s, j, k)$ represents the predicted performance of the s th subject in the k th block during the j th session in the t th task. $\alpha(t)$ denotes the initial performance of the general learning component for the t th task. $\tau(t, s)$ represents the learning rate of the general learning component for the s th subject in the t th task. $s1(t, j)$ represents the magnitude of between-session gain (positive value) or forgetting (negative value) that occurs between the $(j-1)$ th session and the j th session in the t th task. $s2(t, j)$ and $\tau_{s2}(t, j)$ represent the magnitude and rate of within-session relearning during the j th session for the t th task. We kept $s1(t, j)$, $s2(t, j)$ and $\tau_{s2}(t, j)$ unchanged across sessions, in alignment with the previous study which showed that these three components were constant across sessions. In addition, $s3(t, j)$ and $\tau_{s3}(t, j)$ represent the magnitude and rate of within-session adaptation during the j th session for the t th task, specific to the contrast detection task. The magnitude of adaptation varied across sessions, while the rate remained constant.

In summary, the comprehensive model encompassed a total of 71 parameters for the general learning component (including 7 parameters for initial performance in each task, 64 for learning rate, which included 49 subject factors, 7 task factors, and 7 task-specific initial performance factors, along with 1 constant) and 25 parameters for the 4 short-term components (including 14 parameters related to within-session relearning (7 parameters for the magnitude and 7 for the rate in each task), 3 parameters for between-session forgetting (Vernier, Face, and Auditory), 2 parameters for between-session off-line gain (Shape and Contrast), and 6 parameters for within-session adaptation (5 parameters for the magnitude in each session and 1 parameter for the rate of adaptation)). The 96-parameter model serves as the most reduced model, assuming that once all factors affecting learning rate and all short-term effects had been accounted for, neither initial performance nor learning rate was affected by sequence *position*.

To assess the influence of the sequence *position*, we introduced 14 additional parameters to the above model (Eq. 10). Among these, 7 parameters represent coefficients of sequence *position* affecting initial performance and 7 represent the coefficients for learning rate in each task in the general learning component. Hence, in the full model,

$$\begin{aligned} \hat{T}(t, s, j, k) &= \lambda1(t) * \text{position}(t, s) * \alpha(t) * (\text{block}(j-1) + k)^{\lambda2(t) * \text{position}(t, s) * \tau(t, s)} \\ &+ s1(t, j) + s2(t, j)(k-1)^{\tau_{s2}(t,j)} - s3(t, j)(k-1)^{\tau_{s3}(t,j)} \\ \tau(t, s) &= \text{task}(t) + \text{subject}(s) + \text{coeff}_{ini}(t) * \alpha(t) \\ \tau(t, s) &= 0, g, g, g, g (\text{for } j = 1 \sim 5) \\ s2(t, j) &= 0, r, r, r, r (\text{for } j = 1 \sim 5) \\ s3(t, j) &= 1, a2, a3, a4, a5 (\text{for } j = 1 \sim 5) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma (\text{for } j = 1 \sim 5) \\ \tau_{s3}(t, j) &= \beta, \beta, \beta, \beta, \beta (\text{for } j = 1 \sim 5) \end{aligned} \quad (11)$$

where $position(t,s)$ denotes the sequence position of the s th participant in the t th task, $\lambda_1(t)$ is the coefficient for initial performance and $\lambda_2(t)$ is the coefficient for learning rate.

Forward stepwise regression

We used a forward regression approach to test the necessity of the 14 parameters related to sequence position. These steps were followed: (1) in the first iteration, each of the 14 parameters was added to the Comprehensive Model and the one that maximized the goodness of fit R^2 was selected; (2) Starting with the Comprehensive Model and the selected parameter from the first iteration, we tested the contribution of the remaining 13 parameters and identified the parameter with the second-largest contribution; (3) This iterative procedure was repeated until all 14 parameters were ranked by their individual contributions on the model's performance. The best-fitting model must be statistically equivalent to the full model and significantly superior to the most reduced model while using the fewest parameters.

Applying the Comprehensive Model to the retention dataset

When analyzing the retention dataset from the subset of 17 subjects, we used their performance from both the 5 training and re-test sessions. To tailor the comprehensive model to this dataset, we added an additional long-term forgetting component to estimate the magnitude of forgetting that could occur between the 5th session and re-test session. In addition, we introduced separate parameters to account for within-session relearning during the re-test session, instead of using the same parameters for describing these effects in the five training sessions. Now the model without the effect of sequence position is:

$$\begin{aligned} \hat{T}(t, \widehat{s}, j, k) &= \alpha(t) * (block(j-1) + k)^{\tau(t,s)} \\ &+ s_1(t, j) + s_2(t, j)(k-1)^{\tau_{s2}(t,j)} - s_3(t, j)(k-1)^{\tau_{s3}(t,j)} - s_4(t) \\ \tau(t, s) &= task(t) + subject(s) + coeff_{im}(t) * \alpha(t) \\ s_1(t, j) &= 0, g, g, g, g \text{ (for } j = 1 \sim 5) \\ s_2(t, j) &= 0, r, r, r, r \text{ (for } j = 1 \sim 6) \\ s_3(t, j) &= a_1, a_2, a_3, a_4, a_5, a_6 \text{ (for } j = 1 \sim 6) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma \text{ (for } j = 1 \sim 6) \\ \tau_{s3}(t, j) &= \beta, \beta, \beta, \beta, \beta, \beta \text{ (for } j = 1 \sim 6) \end{aligned} \quad (12)$$

where $s_4(t)$ represents the magnitude of long-term forgetting between the 5th and retention sessions in the t th task. The independent parameters for within-session short-term effects consist of the magnitude of relearning r' and the rate of relearning γ' .

To examine the influence of sequence position on both the learning process and the extent of forgetting, we introduced 21 additional parameters to the above model (Eq. 12). Among these, seven parameters represented the coefficients of sequence position on initial performance in each task, seven represented the coefficients of sequence position on the learning rate in each task within the general learning component, and the remaining seven represented the coefficients of sequence position on the magnitude of long-term forgetting in each task. Hence, in the full model,

$$\begin{aligned} \hat{T}(t, \widehat{s}, j, k) &= \lambda_1(t) * position(t, s) * \alpha(t) * (block(j-1) + k)^{\lambda_2(t) * position(t, s) * \tau(t, m)} \\ &+ s_1(t, j) + s_2(t, j)(k-1)^{\tau_{s2}(t,j)} - s_3(t, j)(k-1)^{\tau_{s3}(t,j)} \\ &- (1 + \lambda_3(t) * position(t, s)) * s_4(t) \\ \tau(t, s) &= task(t) + subject(s) + coeff_{im}(t) * \alpha(t) \\ s_1(t, j) &= 0, g, g, g, g \text{ (for } j = 1 \sim 5) \\ s_2(t, j) &= 0, r, r, r, r \text{ (for } j = 1 \sim 6) \\ s_3(t, j) &= a_1, a_2, a_3, a_4, a_5, a' \text{ (for } j = 1 \sim 6) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma \text{ (for } j = 1 \sim 6) \\ \tau_{s3}(t, j) &= \beta, \beta, \beta, \beta, \beta, \beta' \text{ (for } j = 1 \sim 6) \end{aligned} \quad (13)$$

where $position(t,s)$ denotes the position of sequence of the s th subject in the t th task, $\lambda_1(t)$ is the coefficient of the influence of sequence position on initial performance, $\lambda_2(t)$ is the coefficient of the influence of sequence position on learning rate within the general learning component, and $\lambda_3(t)$ is the coefficient of the influence of sequence position on the magnitude of forgetting of the long-term forgetting component. Similarly, we used a forward regression approach to test the necessity of the 21 parameters related to sequence position.

Applying the Comprehensive Model to the dual-task experiment

For the dual-task experiment, the comprehensive model encompassed a total of 15 parameters, including 9 parameters for the general learning component (2 parameters for initial performance in each task, 7 for learning rate consisting of 2 group factors, 2 task factors, and 2 task-specific initial performance factors, along with 1 constant) and 6 parameters related to between-session effect (2 parameters) and within-session relearning components (4 parameters) in the two tasks. The model without the effect of sequence position is:

$$\begin{aligned} \hat{T}(t, \widehat{m}, j, k) &= \alpha(t) * (block(j-1) + k)^{\tau(t,m)} + s_1(t, j) + s_2(t, j)(k-1)^{\tau_{s2}(t,j)} \\ \tau(t, s) &= task(t) + group(m) + coeff_{im}(t) * \alpha(t) \\ s_1(t, j) &= 0, g, g, g, g \text{ (for } j = 1 \sim 5) \\ s_2(t, j) &= 0, r, r, r, r \text{ (for } j = 1 \sim 5) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma \text{ (for } j = 1 \sim 5) \end{aligned} \quad (14)$$

where t represents for the t th task ($t = 1, 2, \dots, 7$), m represents the m th group ($m = 1, 2$), j represents the j th session ($j = 1, 2, \dots, 5$), k represents the k th block in the current session ($k = 1, 2, \dots, 8$), and $block$ represents the number of blocks in each session ($block = 8$). $\hat{T}(t, \widehat{m}, j, k)$ represents the predicted performance of the m th group of subjects in the k th block during the j th session in the t th task. $\alpha(t)$ denotes initial performance of the general learning component for the t th task, and $\tau(t, m)$ represents the learning rate of the general learning component for the m th group at the t th task.

To evaluate the influence of sequence position, we also used the forward regression approach to test the necessity of $\lambda_1(t)$ and $\lambda_2(t)$. The full model can be written as:

$$\begin{aligned} \hat{T}(t, \widehat{m}, j, k) &= \lambda_1(t) * position(t, s) * \alpha(t) * (block(j-1) + k)^{\lambda_2(t) * position(t, s) * \tau(t, m)} \\ &+ s_1(t, j) + s_2(t, j)(k-1)^{\tau_{s2}(t,j)} \\ \tau(t, s) &= task(t) + group(m) + coeff_{im}(t) * \alpha(t) \\ s_1(t, j) &= 0, g, g, g, g \text{ (for } j = 1 \sim 5) \\ s_2(t, j) &= 0, r, r, r, r \text{ (for } j = 1 \sim 5) \\ \tau_{s2}(t, j) &= 0, \gamma, \gamma, \gamma, \gamma \text{ (for } j = 1 \sim 5) \end{aligned} \quad (15)$$

Data availability

The behavior data reported in this study is publicly available in the study's Open Science Framework repository (<https://osf.io/t6jxb/>). Data from Experiment 1A and Experiment 2 has been used to investigate (1) individual differences (<https://osf.io/dgqxv/>), (2) the functional forms of learning curves (<https://osf.io/sbnyr/>). No part of the analyses reported in this study was preregistered before the research was conducted.

Code availability

The code used for the current analysis is publicly available in the study's Open Science Framework repository (<https://osf.io/t6jxb/>). All statistical analyses within this manuscript were conducted in MATLAB (MATLAB 2022a).

Received: 16 December 2024; Accepted: 9 April 2025;
Published online: 09 May 2025

References

- Dosher, B. & Lu, Z.-L. *Perceptual Learning: How Experience Shapes Visual Perception* (The MIT Press, 2020).
- Peter, S. et al. Neuroanatomical disposition, natural development, and training-induced plasticity of the human auditory system from childhood to adulthood: a 12-year study in musicians and nonmusicians. *J. Neurosci.* **43**, 6430 (2023).
- Frank, S. M. et al. Fundamental differences in visual perceptual learning between children and adults. *Curr. Biol.* **31**, 427–432.e425 (2021).
- Sagi, D. Perceptual learning in vision research. *Vis. Res.* **51**, 1552–1566 (2011).
- Hoffman, R. R. et al. *Accelerated Expertise: Training for High Proficiency in a Complex World* (Psychology Press, 2014).
- Seitz, A. R. Perceptual learning. *Curr. Biol.* **27**, R631–R636 (2017).
- Frank, S. M. et al. Supervised learning occurs in visual perceptual learning of complex natural images. *Curr. Biol.* **30**, 2995–3000.e2993 (2020).
- Levi, D. M. Rethinking amblyopia 2020. *Vis. Res.* **176**, 118–129 (2020).
- Fu, Q.-J. & Galvin, J. J. Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif.* **11**, 193–205 (2007).
- Seitz, A. R. Perceptual Learning: changes across the Lifespan. *Curr. Biol.* **31**, R69–R72 (2021).
- Lindenberger, U. & Lövdén, M. Brain plasticity in human lifespan development: the exploration–selection–refinement model. *Annu. Rev. Dev. Psychol.* **1**, 197–222 (2019).
- Fahle, M. & Morgan, M. No transfer of perceptual learning between similar stimuli in the same retinal position. *Curr. Biol.* **6**, 292–297 (1996).
- Fahle, M. Specificity of learning curvature, orientation, and vernier discriminations. *Vis. Res.* **37**, 1885–1895 (1997).
- Li, W., Piech, V. & Gilbert, C. D. Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* **7**, 651–657 (2004).
- Huang, C. B., Lu, Z. L. & Dosher, B. A. Co-learning analysis of two perceptual learning tasks with identical input stimuli supports the reweighting hypothesis. *Vis. Res.* **61**, 25–32 (2012).
- Xiao, L. Q. et al. Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.* **18**, 1922–1926 (2008).
- Kattner, F., Cochran, A., Cox, C. R., Gorman, T. E. & Green, C. S. Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Curr. Biol.* **27**, 840–846 (2017).
- Bavelier, D., Green, C. S., Pouget, A. & Schrater, P. Brain plasticity through the life span: learning to learn and action video games. *Annu. Rev. Neurosci.* **35**, 391–416 (2012).
- Shibata, K., et al. Overlearning hyper stabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nat. Neurosci.* **20**, 470 (2017).
- Seitz, A. R. et al. Task-specific disruption of perceptual learning. *Proc. Natl. Acad. Sci. USA* **102**, 14895–14900 (2005).
- Kirkpatrick, J. et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* **114**, 3521–3526 (2017).
- Tadros, T., Krishnan, G. P., Ramyaa, R. & Bazhenov, M. Sleep-like unsupervised replay reduces catastrophic forgetting in artificial neural networks. *Nat. Commun.* **13**, 7742 (2022).
- Ning, R. & Wright, B. A. Evidence that anterograde learning interference depends on the stage of learning of the interferer: blocked versus interleaved training. *Learn. Mem.* **30**, 101–109 (2023).
- Jeter, P. E., Dosher, B. A., Liu, S. H. & Lu, Z. L. Specificity of perceptual learning increases with increased training. *Vis. Res.* **50**, 1928–1940 (2010).
- Huang, Z., Niu, Z. & Li, S. Reactivation-induced memory integration prevents proactive interference in perceptual learning. *J. Vis.* **23**, 1 (2023).
- Hung, S. C. & Seitz, A. R. Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *J. Neurosci.* **34**, 8423–8431 (2014).
- Wang, R., Cong, L. J. & Yu, C. The classical TDT perceptual learning is mostly temporal learning. *J. Vis.* **13**, <https://doi.org/10.1167/13.5.9> (2013).
- Bejjanki, V. R. et al. Action video game play facilitates the development of better perceptual templates. *Proc. Natl. Acad. Sci. USA* **111**, 16961–16966 (2014).
- Saffell, T. & Matthews, N. Task-specific perceptual learning on speed and direction discrimination. *Vis. Res.* **43**, 1365–1374 (2003).
- Pavlovskaya, M. & Hochstein, S. Perceptual learning transfer between hemispheres and tasks for easy and hard feature search conditions. *J. Vis.* **11**, 8 (2011).
- Amitay, S., Zhang, Y. X. & Moore, D. R. Asymmetric transfer of auditory perceptual learning. *Front. Psychol.* **3**, 508 (2012).
- McGovern, D. P., Webb, B. S. & Peirce, J. W. Transfer of perceptual learning between different visual tasks. *J. Vis.* **12**, 4 (2012).
- Wang, X., Zhou, Y. & Liu, Z. Transfer in motion perceptual learning depends on the difficulty of the training task. *J. Vis.* **13**, 5 (2013).
- Green, C. S., Kattner, F., Siegel, M. H., Kersten, D. & Schrater, P. R. Differences in perceptual learning transfer as a function of training task. *J. Vis.* **15**, 5–5 (2015).
- Yang, J. et al. General learning ability in perceptual learning. *Proc. Natl. Acad. Sci. USA* **117**, 19092–19100 (2020).
- Yang, J. et al. Identifying long- and short-term processes in perceptual learning. *Psychol. Sci.* **33**, 830–843 (2022).
- Bang, J. W., Milton, D., Sasaki, Y., Watanabe, T. & Rahnev, D. Post-training TMS abolishes performance improvement and releases future learning from interference. *Commun. Biol.* **2**, 320 (2019).
- Brashers-Krug, T., Shadmehr, R. & Bizzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).
- Bang, J. W. et al. Consolidation and reconsolidation share behavioural and neurochemical mechanisms. *Nat. Hum. Behav.* **2**, 507–513 (2018).
- Herszage, J. & Censor, N. Modulation of learning and memory: a shared framework for interference and generalization. *Neuroscience* **392**, 270–280 (2018).
- Lu, Z. L. & Dosher, B. A. Current directions in visual perceptual learning. *Nat. Rev. Psychol.* **1**, 654–668 (2022).
- Astorga, G. et al. Adaptive processing and perceptual learning in visual cortical areas V1 and V4. *Proc. Natl. Acad. Sci. USA* **119**, e2213080119 (2022).
- Sigman, M. et al. Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron* **46**, 823–835 (2005).
- Dosher, B. A., Jeter, P., Liu, J. & Lu, Z. L. An integrated reweighting theory of perceptual learning. *Proc. Natl. Acad. Sci. USA* **110**, 13678–13683 (2013).
- Sotiropoulos, G., Seitz, A. R. & Series, P. Performance-monitoring integrated reweighting model of perceptual learning. *Vis. Res.* **152**, 17–39 (2018).
- Dodge, S. & Karam, L. A study and comparison of human and deep learning recognition performance under visual distortions. In: *Proc. 26th International Conference on Computer Communication and Networks (ICCCN)*, Vancouver, BC, Canada, pp. 1–7, <https://doi.org/10.1109/ICCCN.2017.8038465> (2017).
- van de Ven, G. M., Siegelmann, H. T. & Tóliás, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* **11**, 4069 (2020).
- Perkonig, M. et al. Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging. *Nat. Commun.* **12**, 5678 (2021).
- Laborieux, A., Emoult, M., Hirtzlin, T. & Querlioz, D. Synaptic metaplasticity in binarized neural networks. *Nat. Commun.* **12**, 2549 (2021).

50. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* **10**, 647–658 (2009).
51. Yotsumoto, Y., Watanabe, T. & Sasaki, Y. Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron* **57**, 827–833 (2008).
52. Ito, T. & Murray, J. D. Multitask representations in the human cortex transform along a sensory-to-motor hierarchy. *Nat. Neurosci.* **26**, 306–315 (2023).
53. Li, X., Lu, Z.-L., Xu, P., Jin, J. & Zhou, Y. Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *J. Neurosci. Methods* **130**, 9–18 (2003).
54. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
55. Brainard, D. H. Psychophysics software for use with MATLAB. *Spat. Vis.* **10**, 433–436 (1997).
56. Jaeggi, S. M., Buschkuhl, M., Jonides, J. & Perrig, W. J. Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. USA* **105**, 6829–6833 (2008).
57. Levitt, H. Transformed up-down methods in psychoacoustics. *J. Acoustical Soc. Am.* **49**, 467–477 (1971).
58. Fang, F. & He, S. Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron* **45**, 793–800 (2005).
59. Sigman, M. & Gilbert, C. D. Learning to find a shape. *Nat. Neurosci.* **3**, 264–269 (2000).
60. Jaeggi, S. M., Buschkuhl, M., Jonides, J. & Perrig, W. J. Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci.* **105**, 6829 (2008).
61. Myung, I. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **47**, 90–100 (2003).
62. Doshier, B. A. & Lu, Z.-L. The functional form of performance improvements in perceptual learning: learning rates and transfer. *Psychol. Sci.* **18**, 531–539 (2007).

Acknowledgements

This work was supported by National Science and Technology Innovation 2030 Major Projects 2022ZD0204800, National Key Research and Development Program of China 2023YFC3604100, National Natural Science Foundation of China Grants NSFC 32071056, National Natural Science Foundation of China Grants NSFC 32100864, Chinese Academy of Sciences Grant ZDKYYQ20200005 to C.-B. H., National Eye Institute Grant EY017491 to Z.-L. L., Natural Science Foundation of China Grant 32200857, China Postdoctoral Science Foundation Grants 2023M740125, China Postdoctoral Science Foundation Grants 2022T150021 to J.Y.

Author contributions

J.Y., C.B.H., Z.L.L., F.F.Y. contributed to the development of the project. J.Y., J.X., and Z.W. participated in data collection, and J.Y., X.Y., T.W., and Q.M. analyzed the data. J.Y., C.B.H., and Z.L.L. wrote the paper. All authors reviewed the manuscript.

Competing interests

Z.-L.L. holds intellectual property interests in visual function measurement and rehabilitation technologies, and equity interests in Adaptive Sensory Technology, Inc. (San Diego, CA, USA) and Jiangsu Juehua Medical Technology, Ltd (Jiangsu, China). C.-B.H. holds intellectual property interests in visual rehabilitation technologies and equity interests in Jiangsu Juehua Medical Technology, Ltd (Jiangsu, China). All the other authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41539-025-00312-7>.

Correspondence and requests for materials should be addressed to Zhong-Lin Lu or Chang-Bing Huang.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025