Title
Transfer of visual perceptual learning over a task-irrelevant feature through feature-invariant representations: Behavioral experiments and model simulations.

Journal
Journal of Vision, 24(6)

Authors
Liu, Jiajuan
Lu, Zhong-Lin
Dosher, Barbara

Publication Date
2024-06-03

DOI
10.1167/jov.24.6.17

Peer reviewed

Transfer of visual perceptual learning over a task-irrelevant feature through feature-invariant representations: Behavioral experiments and model simulations

Jiajuan Liu
Department of Cognitive Sciences, University of California, Irvine, CA, USA

Zhong-Lin Lu
Division of Arts and Sciences, NYU Shanghai, Shanghai, China
Center for Neural Sciences and Department of Psychology, New York University, New York, NY, USA
NYU-ECNU Institute of Brain and Cognitive Science, Shanghai, China

Barbara Dosher
Department of Cognitive Sciences, University of California, Irvine, CA, USA

A large body of literature has examined specificity and transfer of perceptual learning, suggesting a complex picture. Here, we distinguish between transfer over variations in a "task-relevant" feature (e.g., transfer of a learned orientation task to a different reference orientation) and transfer over a "task-irrelevant" feature (e.g., transfer of a learned orientation task to a different retinal location or different spatial frequency), and we focus on the mechanism for the latter. Experimentally, we assessed whether learning a judgment of one feature (such as orientation) using one value of an irrelevant feature (e.g., spatial frequency) transfers to another value of the irrelevant feature. Experiment 1 examined whether learning in eight-alternative orientation identification with one or multiple spatial frequencies transfers to stimuli at five different spatial frequencies. Experiment 2 paralleled Experiment 1, examining whether learning in eight-alternative spatial-frequency identification at one or multiple orientations transfers to stimuli with five different orientations. Training the orientation task with a single spatial frequency transferred widely to all other spatial frequencies, with a tendency to specificity when training with the highest spatial frequency. Training the spatial frequency task fully transferred across all orientations. Computationally, we extended the identification integrated reweighting theory (I-IRT) to account for the transfer data (Dosher, Liu, & Lu, 2023; Liu, Dosher, & Lu, 2023). Just as location-invariant representations in the original IRT explain transfer over retinal locations, incorporating feature-invariant representations effectively accounted for the observed transfer. Taken together, we suggest that feature-invariant representations can account for transfer of learning over a "task-irrelevant" feature.

Introduction

The performance in visual perceptual tasks is often highly dependent on experience, with performance improvements arising from perceptual learning. When learning one task improves performance on another untrained one, the learning transfers or generalizes; whereas, if it does not, then the learning is deemed specific. The circumstances in which visual perceptual learning occurs and whether and how learning in one task transfers to another are major theoretical issues that have been extensively researched. Specificity to the trained judgment has been noted for many stimulus features, including cases of specificity to retinal training location (Karni & Sagi, 1991). Other research has investigated training protocols that might increase transfer (for reviews of perceptual learning and/or transfer, see Connolly, 2019; Dosher & Lu, 2020; Fine & Jacobs, 2002; Green, Banai, Lu, & Bavelier, 2018; Lu & Dosher, 2022; Sagi, 2011).

The focus of this paper is a relatively unstudied aspect of transfer: namely, whether perceptual learning transfers over variations in task-irrelevant stimulus variations. Or, to put it differently, whether training

Citation: Liu, J., Lu, Z.-L., & Dosher, B. (2024). Transfer of visual perceptual learning over a task-irrelevant feature through feature-invariant representations: Behavioral experiments and model simulations. *Journal of Vision*, *24*(6):17, 1–24, https://doi.org/10.1167/jov.24.6.17.

https://doi.org/10.1167/jov.24.6.17 Received October 30, 2023; published June 25, 2024 ISSN 1534-7362

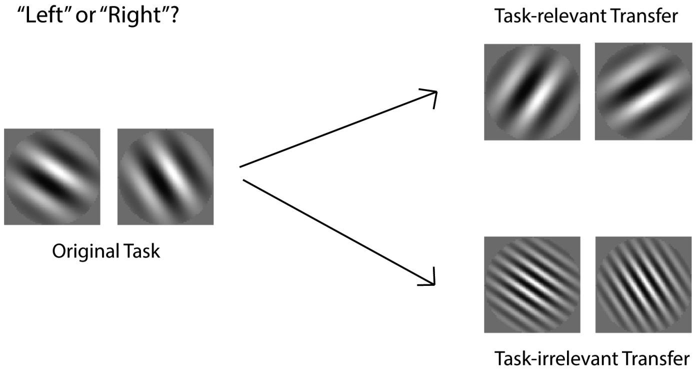

Figure 1. Examples of transfer within a task-relevant dimension (top; to an orthogonal reference angle in an orientation identification task) and within a task-irrelevant dimension (bottom; to a different spatial frequency in the same orientation identification task).

generalizes to stimuli that look different due to changes in a task-irrelevant feature of the stimulus. For example, if observers are trained to judge orientation for a high-spatial-frequency stimulus, does the ability to judge orientation transfer to stimuli with other spatial frequencies? Alternatively, if observers are switched to judge a different set of orientations, this requires transfer to new stimuli in the task-relevant dimension. Generalization of learning has a practical value—it is much less efficient if every possible combination of trained features must be learned. On the other hand, visual perceptual learning may capitalize on all of the available information in the exact stimuli that are trained to optimize performance on the trained task, leading to specificity. We aimed to test these possibilities. Examples of transfer within the task-relevant dimension and a task-irrelevant dimension are shown in Figure 1 for a two-alternative task. Testing multiple graded changes in the task-irrelevant dimension can be used to examine transfer tuning functions.

The current project examined the role of task-irrelevant transfer behaviorally in two studies—one in which observers learn to perform multi-alternative orientation absolute identification using one or several spatial-frequency variations in the training stimuli, and another in multi-alternative spatial-frequency identification using one or several orientation variations in the training stimuli. This design measured tuning in transfer after training with a single level of the task-irrelevant dimension. We also extended and tested a model of visual perceptual learning, the identification version of the integrated reweighting theory (I-IRT) (Dosher, Jeter, Liu, & Lu, 2013: Dosher, Liu, & Lu, 2023; Liu, Lu, & Dosher, 2023). This extension incorporates feature-invariant representations to account for our new data indicating substantial transfer across task-irrelevant variations in stimulus appearance.

Specificity and transfer

There is an extensive literature on specificity and transfer in visual perceptual learning in many different perceptual tasks. Many early perceptual learning studies showed significant specificity for the trained tasks, such as specificity to the trained orientation (Schoups, Vogels, & Orban, 1995), location (Crist, Kapadia, Westheimer, & Gilbert, 1997), motion direction (Ball & Sekuler, 1987), or the eye trained (Karni & Sagi, 1991). Several theoretically influential examples involved specificity to the retinal location of training, leading researchers to hypothesize early visual cortical areas as a substrate for learning (Karni & Sagi, 1991).

In many cases, however, tasks exhibit both some specificity and some transfer to the new task variants (Jeter, Dosher, Liu, & Lu, 2010; for a review, see Dosher & Lu, 2020; Lu & Dosher, 2022). Even testing the same stimuli, different tasks or judgments can yield both specificity and transfer (Green, Kattner, Siegel, Kersten, & Schrater, 2015). Transfer between tasks is asymmetric in some cases (Dosher & Lu, 2005; Ahissar & Hochestein, 1997; Jeter, Dosher, Petrov, & Lu, 2009; Liu & Weinshall, 2000). In addition, a substantial literature has investigated "double training" in more than one task (e.g., orientation task 1 in location 1 and orientation task 2 in location 2) and improved transfer across locations and tasks (Hu, Wen, Chen, & Yu, 2021; Wang et al., 2016; Xiao et al., 2008; Xiong, Zhang, & Yu, 2016; Zhang et al., 2010). From all these informative yet nuanced results, one conclusion holds: Details matter. The specific stimulus, task, training sequence, and measurements may all affect the level of transfer.

Task-relevant and task-irrelevant stimulus feature transfer

We suggest an added classification of transfer into two categories: transfer over variation in a *task-relevant* stimulus feature or over variation in a *task-irrelevant* stimulus feature. This may help disentangle the learning transfer literature, as these two types of transfer may require different underlying mechanisms. This project focuses on the little-studied case of transfer over altered task-irrelevant stimulus features.

An example of transfer (or *specificity*) over a task-relevant feature is training an orientation identification task centered around one reference orientation that transfers to an orientation identification task centered around a different reference orientation. Many of the classic examples of specificity are of this kind, such as for orientation (Dosher & Lu, 2007; Schoups et al., 1995; Vogels & Orban, 1985) or motion direction (Ball & Sekuler, 1987; Watanabe et al., 2002). In these cases, the observer's judgments reflect activity in what are likely to be distinct neural representations in the training and transfer tasks (e.g., around the horizontal and vertical orientations).

The best-known example of task-irrelevant transfer occurs when visual perceptual learning transfers to a different retinal location. The stimuli and judgments are identical, but they are tested in another location relative to fixation. One well-known example demonstrated specificity of training to location in the texture discrimination task (Karni & Sagi, 1991). Others involved orientation tasks in different retinal locations (Schoups et al., 1995). In some cases, a mixture of transfer and specificity is observed (e.g., Dosher et al., 2013; Schoups et al., 1995; Shiu & Pashler, 1992), whereas in others the specificity is nearly complete (e.g., Ball & Sekuler, 1987). Another such feature is specificity (Karni & Sagi, 1991), or generalization over the eye of training (Schoups et al., 1995). External noise level is also an example of transfer over a task-irrelevant feature, which yields asymmetric transfer (Dosher & Lu, 2005). One closely relevant study examined transfer over task-relevant and task-irrelevant features in the same task: switching only orientations, only retinal locations, or both in an orientation identification task (identifying a stimulus orientation as clockwise or counterclockwise $\pm 5°$ of one or another diagonal reference angle) (Dosher et al., 2013). A mix of specificity and transfer occurred in all cases, with more specificity for changes in orientation alone (task-relevant transfer), and the most transfer for changes in location alone (task-irrelevant transfer).

The current project examined transfer over a task-irrelevant feature other than location—one that alters stimulus appearance. It trains orientation judgments with stimuli of one spatial frequency and assesses transfer of the same orientation judgment to stimuli of another spatial frequency (as in Figure 1), or it trains spatial frequency judgments with stimuli at one orientation and tests transfer to stimuli of other orientations. There are few exact predecessors of this in the perceptual learning literature. Perhaps the closest is training contrast detection, which leads to graded transfer near the spatial frequency of the training stimulus (Hou et al., 2011; Huang, Zhou, & Lu, 2008; Sowden, Rose, & Davies, 2002; Zhou et al., 2006).

There are theoretical arguments supporting both specificity and transfer across variation in a task-irrelevant feature. Learning a visual task should take advantage of full knowledge of the stimulus (perhaps by learning image templates that match the trained stimuli). This might optimize performance on a specific trained task but lead to more specificity. However, from a practical perspective, transfer of training across a task-irrelevant feature improves performance over many task variants when training on only one task-irrelevant feature value. This transfer could be accomplished by pooling over the task-irrelevant feature to combine inputs from multiple feature-dependent representations. Another claim (Raviv, Lupyan, & Green, 2022) is that varying the stimuli trained in a task increases transfer. This agrees with using variation in a task-irrelevant feature during training to encourage related generalization (Manenti, Dizaji, & Schwiedrzik, 2023). The idea of variation during training is a point we return to later.

Related results on pooling

Pooling over task-irrelevant dimensions has been used to explain the results in some psychophysical studies. In a series of experiments, Olzak and colleagues (Olzak & Thomas, 1992; Olzak & Thomas, 1999; Olzak & Wickens, 1997; Thomas & Olzak, 1996) studied masking and cue combination in discriminating spatial frequency or orientation using complex gratings. For example, in orientation identification, the information from two widely different spatial frequency components seems to be integrated, so that performance for a congruent stimulus (in which orientation information is of the same sign in both spatial frequency components) is always much better than that of an incongruent stimulus (in which the orientation judgment conflicts for two spatial frequency components). Similar results also hold for spatial frequency judgments from complex gratings with both horizontal and vertical orientation components.

To account for these results, Olzak et al. proposed summation/pooling across the spatial frequency channels ("cigar" model) for orientation judgments and across orientation channels ("doughnut" model) for spatial frequency judgments. The "cigar" model has processing stages that are broadly tuned for spatial frequency but narrowly for orientation, and the "doughnut" has processing stages that are broadly tuned for orientation but narrowly for spatial frequency. The pooling in these studies occurred for simultaneously available components of complex stimuli, whereas relevant pooling in the perceptual learning studies here occurred through activation in different trials of corresponding pooled representations that we refer to as (task-irrelevant) feature invariant. The idea of related pooled representations also has roots in neurophysiological studies, a point considered in the Discussion.

Integrated reweighting theory and transfer

The integrated reweighting theory (IRT) provides a theoretical and computational framework for both learning and transfer in visual tasks. In the

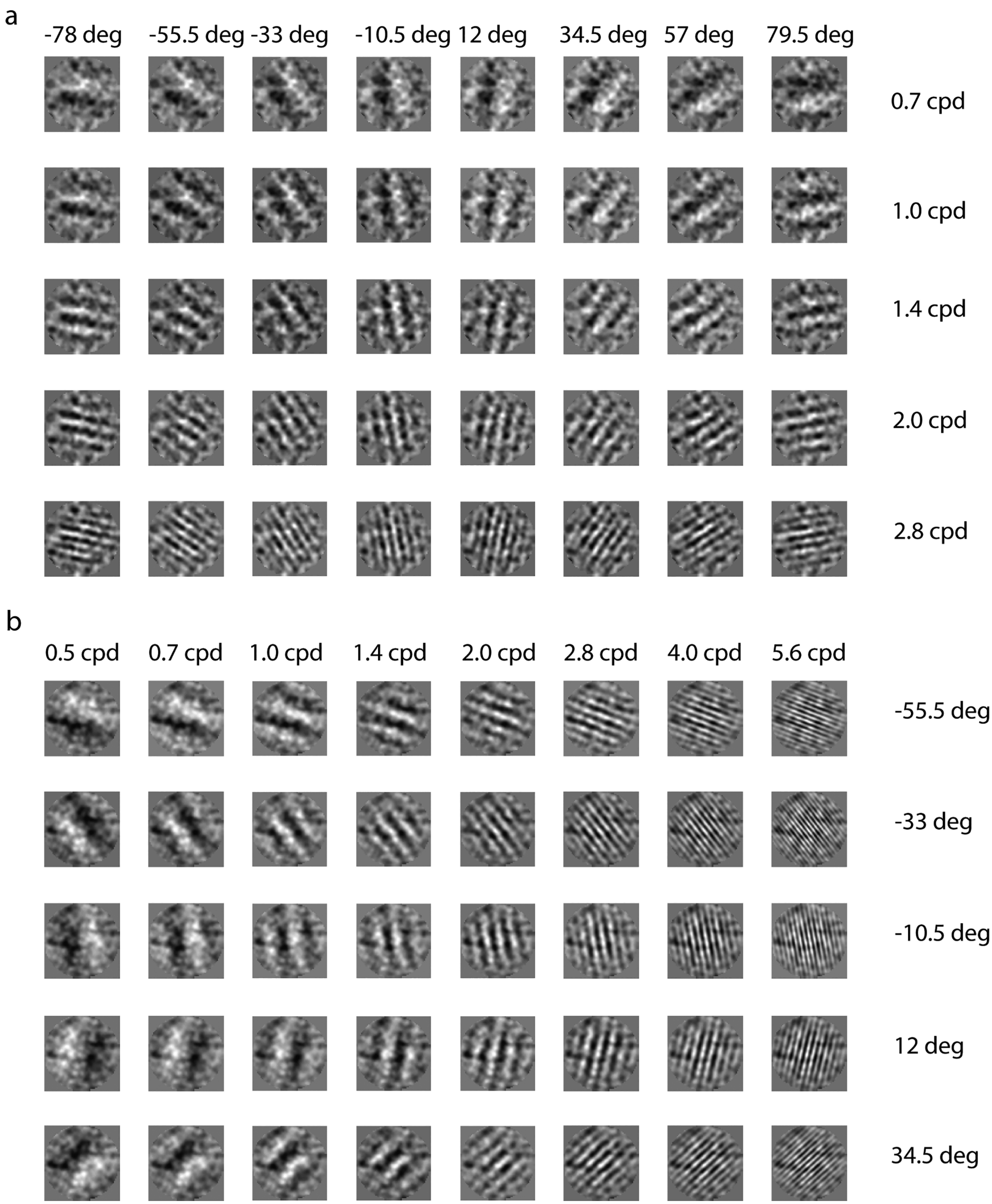

Figure 2. Stimuli of the two experiments were Gabors displayed with external noise. (**a**) Gabor stimuli for the eight-alternative forced choice (8AFC) orientation-identification task. Every stimulus is of one of the eight possible orientations and one of the five possible spatial frequencies. (**b**) Gabor stimuli for the 8AFC spatial-frequency task. Every stimulus is of one of the eight possible spatial frequencies and one of the five possible orientations.

reweighting models, the most relevant and active sensory representation units are increasingly weighted to drive activation in a decision unit, whereas less relevant representations are downweighted; learning is embodied in the process of reweighting. The original reweighting model has banks of representation units jointly tuned to orientations and spatial frequencies that are location specific (Petrov, Dosher, & Lu, 2005; Petrov, Dosher, & Lu, 2006). The IRT also includes banks of orientation and spatial frequency tuned units that are location invariant—units that respond to stimuli in all retinal locations—to account for location transfer (Dosher et al., 2013). Transfer over location is mediated by learning of weights from location-invariant representations to decision, in addition to location-specific representations whose weights reflect learning that is specific to location. The IRT accounts for the different amounts of transfer and specificity when only orientation, only location, or both are switched from training to transfer in orientation discrimination (Dosher et al., 2013). The same representation structure is also used to account for effects of double training of multiple tasks in multiple locations, which (as reviewed earlier) often produces strong transfer (Talluri, Hung, Seitz, & Series, 2015), and the interactions of training in multiple locations in a task roving paradigm (Dosher, Liu, Chu, & Lu, 2020).

The current project examines whether transfer over task-irrelevant features occurs empirically in multi-alternative identification tasks or, alternatively, if there is significant specificity to the training stimulus, or some of both. By analogy to the location-invariant representations, feature-invariant representations that pool over task-irrelevant feature variation might account for task-irrelevant transfer to untrained stimuli. For example, orientation-tuned representations that pool over spatial frequencies may enable transfer of orientation identification trained in one spatial frequency to others, or spatial-frequency tuned representations that pool over orientations can enable transfer of spatial-frequency identification trained in one orientation to others. We test whether this extended identification IRT (eI-IRT) can account for the mixture of transfer and/or specificity demonstrated in the behavioral data. The computational model and its fit to the data are considered in "The extended I-IRT" section following the reporting of behavioral results.

Design and rationale

Our experiments measured the transfer of perceptual learning over task-irrelevant features by comparing post-training performance to pre-training performance in multi-alternative absolute identification. Improvements reflect learning when the test stimulus matched the trained one and transfer when the test stimulus differed from the training stimulus in task-irrelevant feature. Pre-training measurements accounted for any differences in performance for different values of the task-irrelevant dimension by providing baselines. In each trial of the experiment, the test stimulus occurred in one of two peripheral visual locations (e.g., top left, lower right of fixation), pre-cued just slightly before the stimulus, to control fixation. The pre- and post-training assessments used an intermediate contrast condition in high external noise, where performance differences are largest. No feedback was given to limit learning during assessments alone (assessed in a no-training control group).

The task was eight-alternative absolute identification of either orientation (Experiment 1) or spatial frequency (Experiment 2), and the stimulus sets took multiple values of the task-relevant and the task-irrelevant features (see Figure 2). *N*-alternative tasks are especially useful in comparing different transfer conditions, because their low guessing rate (1/8, or 12.5% here) yields more information on each trial, so assessment can be done with fewer trials (Hou, Lesmes, Bex, Dorr, & Lu, 2015). In these studies, we assessed transfer at five values of the task-irrelevant feature, producing a transfer function that, if specificity occurs, should center around the trained task-irrelevant value.

The training in the experimental groups occurred in a design known to produce learning in these eight-alternative identification tasks using full response feedback or supervision (correct response indicated) and three levels of contrast (low, medium, and high) intermixed in multiple sessions (Dosher et al., 2023; Liu et al., 2023). Following the baseline assessment, multiple sessions were trained with one or all values of the task-irrelevant feature (in four different groups) with high external noise, a post-training assessment, two sessions in zero external noise, and a final post-training assessment. Including both zero and high external noise trials better constrained the model estimates of additive internal noise (see Dosher & Lu, 2020). The statistical analysis of the training data is presented in Supplementary Tables S1-2 and S2-2.

Behavioral experiments

Methods

Design

The two experiments had the same design: pre- and post-tests to assess performance improvement

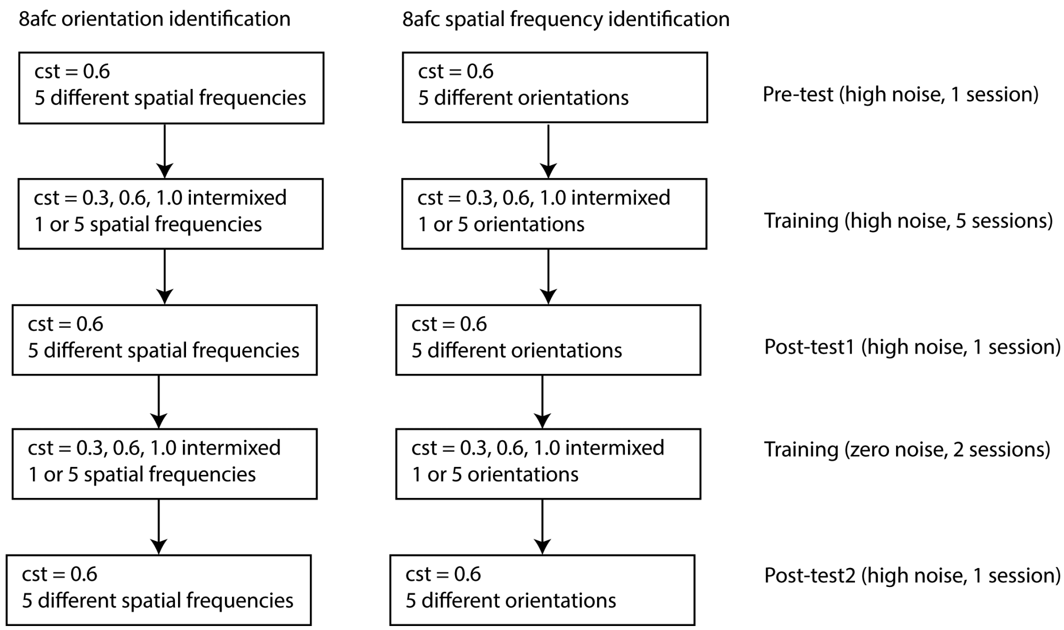

Figure 3. The experimental design in the training conditions. Both experiments had a session of pre-test, five sessions of training with high noise, a post-test, two sessions of training with zero noise, and another post-test. The control group has just the first session and seventh session of pre- and post-tests, with no training in between.

and generalization and training under both high and zero external Gaussian noise between these pre- and post-tests. Both experiments had 10 sessions, where the first, seventh, and 10th sessions were pre- and post-tests. The second to ~sixth sessions included training under high noise, and the eighth and ninth sessions included training under zero noise. There were four training groups in each experiment. Experiment 1 had three training groups trained at one of three spatial frequencies (e.g., low, medium, or high spatial frequency); Experiment 2 included one of three orientations and a fourth that intermixed all task-irrelevant stimuli during training. There was also a control group who participated in the pre- and post-tests only, with a 6-day gap in between to mimic the training group schedule. This allowed evaluating any learning that occurred from the pre- to post-tests alone.

In Experiment 1, observers discriminated the eight orientations of a Gabor patch in one pre-cued location of two possible locations (top left or bottom right), tested over trials with stimuli of one or more spatial frequencies. In pre- and post-tests, equal numbers of stimuli at five different spatial frequencies were tested. The contrast of the stimulus in the pre- and post-tests was 0.6 (intermediate contrast) presented in high noise. No feedback was presented in pre- and post-tests. The training sessions (five sessions in high external noise and two in zero external noise) included stimuli displayed at contrasts of 0.3, 0.6, and 1.0 using a single spatial frequency or a set of spatial frequencies (low, middle, high, or all) in the different groups. Response feedback was provided after each trial during the training sessions (whether the response was correct and the identity of the correct response). A parallel arrangement was used for the eight-alternative spatial-frequency identification in Experiment 2. (See Figure 2 for stimuli and Figure 3 for a diagram of the experimental design.)

Observers

The experiments included 102 observers. All subjects signed a written Institutional Review Board consent from the University of California, Irvine. Observers were required to have normal to corrected-to-normal vision. Each observer in the experimental groups completed the 10-session experiment, spanning roughly 2 to 3 weeks. Control groups completed two sessions of pre- and post-tests, with a 6-day gap between sessions. Each pre- or post-test session tested 800 trials, and each training session had 960 trials. Thus, each observer in training conditions participated in 9120 trials, and each observer in control groups participated in 1600 trials, for a total of 380,800 trials overall in each of the two experiments. Two participants were excluded (one failed to show any learning, and so there was no way to analyze transfer; another misunderstood the task, and the pre-test performance was below chance); this resulted in 10 observers for each group in each experiment.

Stimuli and apparatus

The stimuli were Gabor patterns with an orientation and central spatial frequency. In Experiment 1, the orientation of the stimulus was randomly chosen from eight possible angles (−78°, −55.5°, −33°, −10.5°, 12°, 34.5°, 57°, or 79.5°), with its central spatial frequency at one or more spatial frequencies. In pre- and post-tests, the spatial frequency of the stimulus was randomly chosen on each trial from five possible ones (0.7, 1.0, 1.4, 2.0, or 2.8 cycles per degree [cpd]). In the training sessions between pre- and post-tests, orientation identification was trained in separate groups with stimuli with a designated spatial frequency (low at 0.7 cpd, middle at 1.4 cpd, and high at 2.8 cpd, or randomized over all five spatial frequencies as in pre- and post- tests). Training consisted of five sessions in high external noise and then two sessions in zero external noise. Experiment 2 evaluated learning and transfer in spatial-frequency identification and followed the same design, with the spatial frequency randomly chosen on each trial from eight possible ones (0.5, 0.7, 1.0, 1.4, 2.0, 2.8, 4.0, or 5.6 cpd). In pre- and post-tests, spatial-frequency identification was tested at five orientations, randomized over trials (−55.5°, −33°, −10.5°, 12°, or 34.5°). In training phases, the stimuli used one orientation (left at −55.5°, middle at −10.5°, or right at 34.5°) or all five orientations (as in pre- and post-tests), depending on the group. (See Figure 2 for an example of all stimuli in high noise.) Thus, each experiment included four groups of observers in the various training conditions and one group of observers in the control condition.

On each trial, the Gabor stimulus was displayed either at the top left or bottom right corner of the screen. The 64 × 64 pixel patch is defined by $l(x, y) = l_0(1.0 + c\sin(2\pi f(y\sin(\theta) + x\cos(\theta)) \times e^{\frac{x^2+y^2}{2\sigma^2}})$, where θ and f are the chosen orientation and spatial frequency, respectively; $\sigma = 0.7°$, the standard deviation of the Gaussian envelope; c is the maximum contrast; and l_0 is the mid-gray background luminance. In our experiment, c was one of the three contrasts: 0.3, 0.6, or 1.0.

The external noise images were generated independently for each trial. Each noise image was composed of 2 × 2-pixel noise elements, whose contrasts were randomly chosen from a Gaussian distribution with mean 0 and standard deviation 0.33 (for the orientation task) or 0.25 (for the spatial frequency task) and filtered through a second-order Butterworth bandpass filter (cut-off frequencies at 1.4 cpd and 5.6 cpd, respectively). On each trial, the stimulus sequence was two external noise images followed by one signal image and then followed by two more external noise images (NNSNN). Each external noise image was independently generated. These images were flashed through quickly at the refresh rate, and participants perceived a single noisy Gabor through temporal integration. Stimuli were generated using MATLAB (MathWorks, Natick, MA) with Psychtoolbox 3 on a Dell PC (Dell Technologies, Round Rock, TX) and displayed on a 20-inch monitor (ViewSonic, Brea, CA). The color monitor was set at a refresh rate of 60 Hz and resolution of 640 × 480 pixels and in a pseudo-monochrome with luminance linearized into 127 levels by a visual calibration procedure (Lu & Dosher, 2013). The minimum, maximum, and mid-gray background luminance values were 1, 67, and 34 cd/m^2, respectively. Images were presented at 5.3° visual angle eccentricity, subtended 2.8° × 2.8°. Participants sat 83 cm from the monitor, and a chin rest was used to stabilize the distance.

Procedure

The task was explained to the observers, including showing them printed examples of the stimuli. Then, observers practiced 32 exemplar trials in zero noise to make sure they understood the task. In the actual experiment, in each trial, a central fixation cross and two sets of location marks on the two locations (top left/bottom right) were shown for 300 ms, then the stimulus sequence (NNSNN) was flashed through in the stimulus location. Each noise frame appeared for one refresh count (16.7 ms), and the signal frame appeared for two refresh counts (33.3 ms). A central pre-cue arrow indicating the tested location appeared 100 ms before the first stimulus frame and reappeared as a post-cue after the stimulus display was complete. The signal contrast was randomly set at 0.3, 0.6, or 1.0 (or 0.6 during pre- and post-assessments). Observers pressed one of the eight keys (a, s, d, f, j, k, l, or ;) as the response, one for each orientation (Experiment 1) or spatial frequency (Experiment 2). In training trials, observers received feedback that indicated response accuracy and the correct response.

Data analysis

Learning and transfer were assessed based on performance of pre-and post-tests, as well as in the training sessions. Performances in the two post-tests were essentially equivalent (see Results and Figures 4 and 6), so data from the two post-tests were combined. For learning, we looked at performance improvement over the training sessions, as well as the improvement from pre-test to post-tests compared with a control group with no training sessions.

To assess transfer, we compared performance improvements among different spatial frequencies (Experiment 1) or orientations (Experiment 2) from

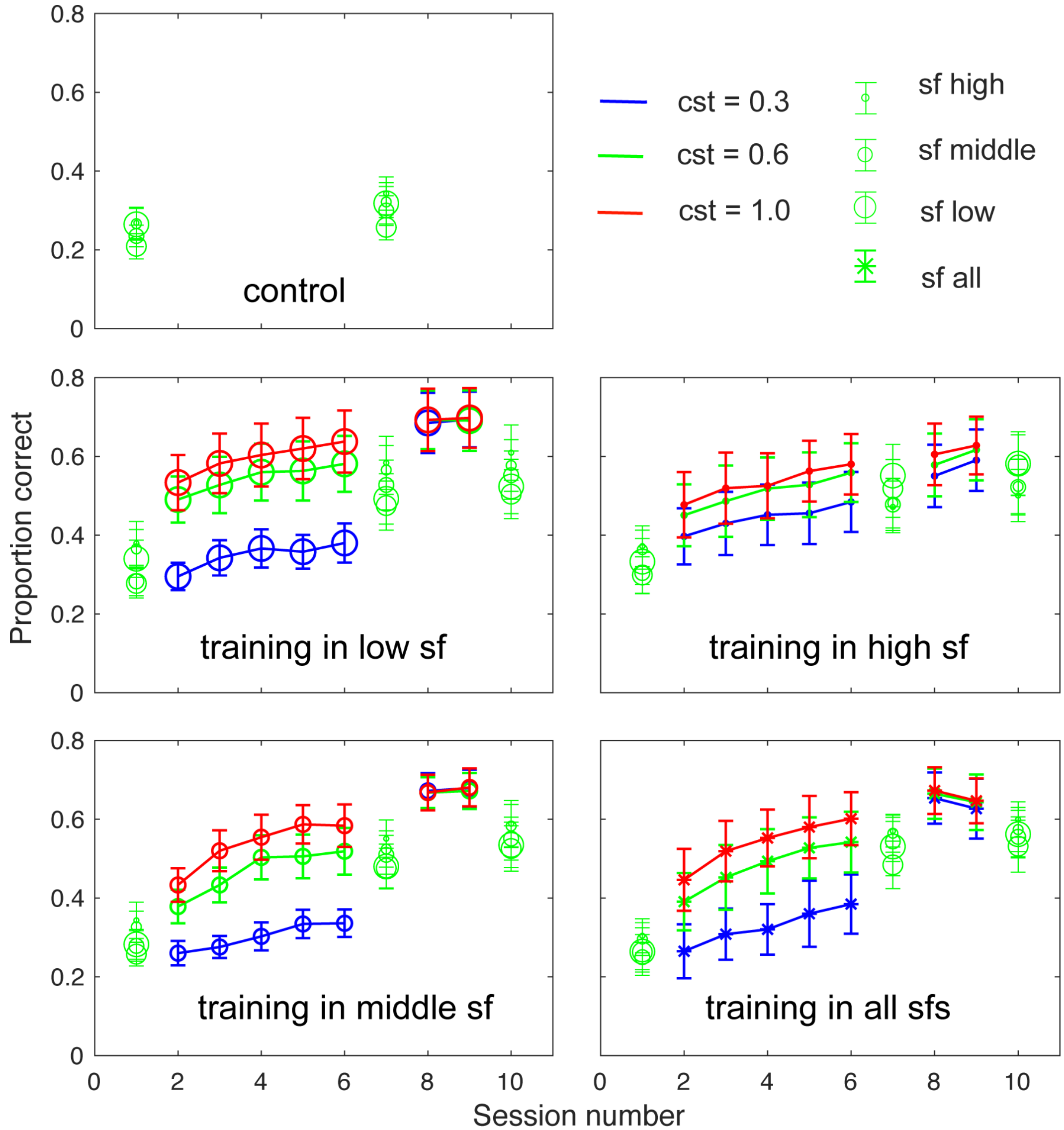

Figure 4. Experimental results of the orientation identification task (Experiment 1). Five panels are the five groups: control (pre- and post-tests only, top left), training with low spatial frequencies (middle left), high spatial frequencies (middle right), middle spatial frequencies (bottom left), and all spatial frequencies (bottom right). All pre- and post-tests were done with all five spatial frequencies with the middle contrast (0.6). Training was done with three contrast levels (0.3, 0.6, and 1.0). The control group changed little from pre- to post-test, whereas the overall post-tests performance of all spatial frequencies improved after training regardless of which spatial frequencies (group) were trained, signifying both learning and transfer (over a task-irrelevant feature).

pre-test to post-test in each of the training groups. We defined an improvement index, I, as

$$I = \left(pc_{post} - pc_{pre}\right) / \left(1 - pc_{pre}\right)$$

The improvement index shows what percentage of the maximum possible improvement was realized by learning. (This index is contrast dependent; however, all pre- and post-tests in this design used the same 0.6 contrast.) Analyses of variance, both traditional and Bayesian, were performed to test whether there was a significant difference between observed improvements as a function of the training group.

Results

Experiment 1: Orientation identification with different spatial frequency

Training the eight-alternative orientation identification task generated significant learning, and most learning also transferred across different spatial frequencies. Figure 4 shows the proportion correct for all sessions, including pre-test, training in high noise, the first post-test, training in zero noise, and the final post-test, for all training groups (low, middle, high, or

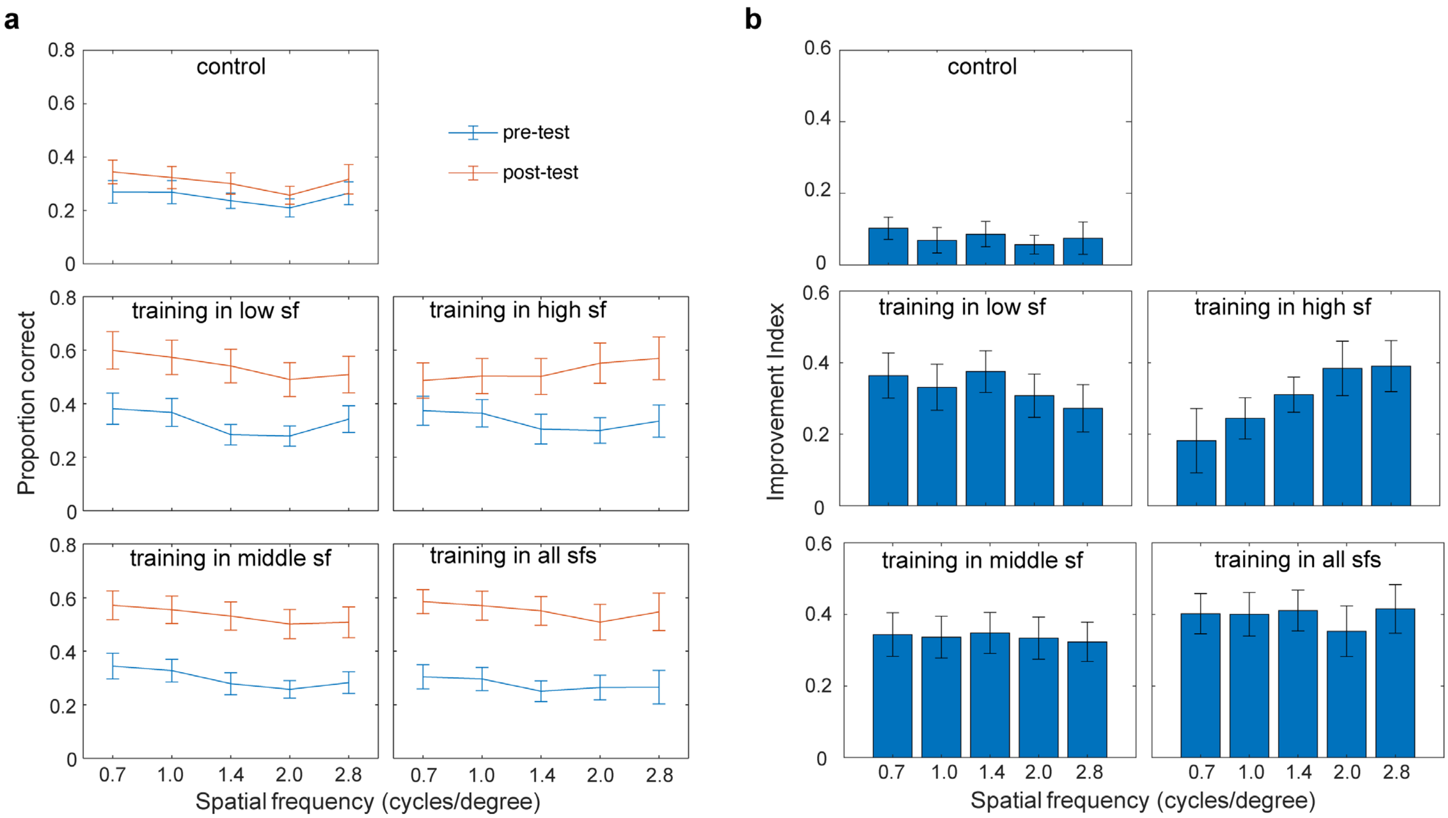

Figure 5. Results from the pre- and post-tests in orientation identification (Experiment 1; see text). (**a**) The proportion correct results. Change for the control group is limited. In training groups, the (near) parallel line of pre- and post-tests represents (almost) complete transfer. Top left shows control; middle left, training in a low spatial frequency; middle right, training in a high spatial frequency; bottom left, training in a middle spatial frequency; and bottom right, training in all five spatial frequencies. (**b**) Pre- to post-tests improvement index (see text). Similar bar length across different spatial frequencies signifies transfer. The error bars are standard error of the mean (*SEM*).

all five values of the task-irrelevant feature) and the control group.

For all of the training groups, both learning due to training and generalization were manifested. Learning can be seen in the higher performance levels in the post-tests compared with those in the pre-test and in the learning curves in the high external noise training sessions. Transfer is seen in improvements that extended to all spatial frequencies in post-tests, whether the orientation task was trained at that spatial frequency or that spatial frequency was untrained (see statistical tests below). The proportion correct was equivalent in the first and second post-tests (most $p > 0.1$; see Supplementary Table S1-1 for details), so the results from the two post-tests were combined in the following analyses. Analyses of variance on pre-test (session 1) performance found no significant difference among groups, $F(3, 36) = 0.394$, $p > 0.7$, before training, and, upon training (sessions 2 to 6), the analyses found no difference in learning among groups, $F(3, 36) = 0.287$, $p > 0.8$. Taken together, then, all training groups started with similar initial performance and showed similar improvement over the training sessions (see detailed analysis of the training data in Supplementary Table S1-2).

Next, we look in more detail at the pre- and post-tests to assess both learning and transfer (see Figure 5a). Performance improvements correspond with the gap between two curves (pre-test and post-test) in each subpanel. Transfer across all task-irrelevant variations in the training stimuli is seen in the near-parallel form of the pre- and post-curves (similar improvement across all spatial frequencies) regardless of the spatial frequency (or frequencies) used to train the orientation identification task (low, medium, high, or all groups). There was almost complete transfer in most cases (training with low or middle spatial frequencies transferred to all other ones; training with high spatial frequency mostly transferred to other spatial frequencies, albeit showing a trend for reduced transfer in the lowest spatial frequencies). Statistical analysis confirmed these observations (see Supplementary Table S1-3 for details).

An improvement index was computed from the pre- and post-training test scores (see Methods) for each group and each spatial frequency (Figure 5b). Again, virtually all spatial frequency tests showed significant improvement in orientation identification from pre- to post-test in all training groups (improvement index > zero, all $ps < 0.01$, with the only exception being for lowest spatial frequency tests in the group trained with high spatial frequency ($p = 0.07$). (See Supplementary Table S1-4 for details.) In comparison, any improvements in the control group from pre-test to the post-test were relatively small (only improvement of the first and third spatial frequencies reached significance and the others did not; see Figure 5b, top-left panel).

Traditional statistical testing evaluates the evidence for an alternative hypothesis, which some have argued does not directly assess evidence in favor of a null hypothesis (Wagenmakers, 2007). Because the theoretical point is the essentially equivalent pattern of transfer in all training groups, we performed Bayesian tests using the Bayesian information criterion (BIC), or fitness of competing models, method (Masson, 2011), to assess the strength of evidence favoring the null hypothesis of equivalent transfer across the task-irrelevant stimulus dimension, as well as the evidence favoring the alternative hypothesis of differential transfer. Traditional p values reflect $p(D/H_0)$, or the probability of observing current data given the null hypothesis. The Bayesian method provides estimates of $p(H_0/D)$ and $p(H_1/D)$, or the probability that the null (or alternative) hypothesis is true given the data.

We carried out a Bayesian analysis on the improvement indices, with trained spatial frequency as a between-group factor and spatial frequency of the test stimulus as a within-group factor. In this analysis, with five groups (including the control), the calculated effect sizes for group, spatial frequency and the interaction between them were 0.330, 0.0181, and 0.194, respectively, and the $p_{BIC}(H_0/D)$ values were 0.00, 1.00, and 1.00, respectively. Values near 1.0 provide very strong evidence that the null hypothesis is true, whereas values near 0 favor the alternative hypothesis. The strong evidence for the alternative hypothesis comparing across groups solely reflected the difference between the control group and all four training groups. When excluding the control group, evidence for group differences was eliminated, with respective effect sizes of 0.0388, 0.0209, and 0.209 for group, spatial frequency, and the interaction term, respectively, and the $p_{BIC}(H_0/D)$ values were 0.99, 1.00, and 1.00, respectively, showing that the degree of learning was not significantly affected by training group, spatial frequency of the stimuli, or the interaction between them. Thus, the Bayesian analysis provided direct evidence consistent with broad transfer across all spatial frequencies in all four training groups. In this case, including all spatial frequencies during training resulted in essentially equivalent post-training results. (The classic frequentist analysis is presented in Supplementary Tables S1–S3.)

We also examined the response confusion matrices (the frequency of different responses given to each stimulus), which showed performance improvements across all orientations in the stimuli (not biasing responses to a subset of stimuli) in the training groups. The control group showed minimal change. See Supplementary Figure S1-1 for a visualization of the heat diagrams of the pre- and post-tests in all five groups and Supplementary Figure S1–S2 for response confusion profiles, which show the confusions with near stimuli.

In summary, then, Bayesian (and frequentist) statistical methods confirmed the null hypothesis of no difference between training groups, leading to the conclusion that training with any of the spatial frequencies transfers essentially completely to others in the orientation identification task, that the amount of transfer was similar after training with all spatial frequencies, and that such generalization is not explained by learning from the pre-test to post-test improvement alone (control conditions). Similar conclusions were supported by an analysis of the pattern of response confusions.

Experiment 2: Spatial-frequency identification task with different orientations

Experiment 2 paralleled Experiment 1 with the spatial-frequency identification task that required identification of eight different spatial frequencies tested using five different stimulus orientations. Here, the four main groups trained on a left-most orientation, a right-most orientation, a middle orientation, or all five orientations. Figure 6 shows the effect of spatial-frequency identification training with stimuli of different orientations, including performance on the pre-test, the training sessions using external noise, the first post-test, training in zero external noise, and the second post-test. Paralleling results of Experiment 1, there was significant learning in all training conditions as measured by improvements from pre- to post-test performance and across training sessions, while the control group showed little change of performance. Here, there appeared to be a complete transfer of learning—the performance in the pre- and post-tests for the groups that trained with stimuli of different orientations was essentially the same. Figure 7 shows nearly parallel lines from pre-test to post-test for all orientations regardless of training group. Once

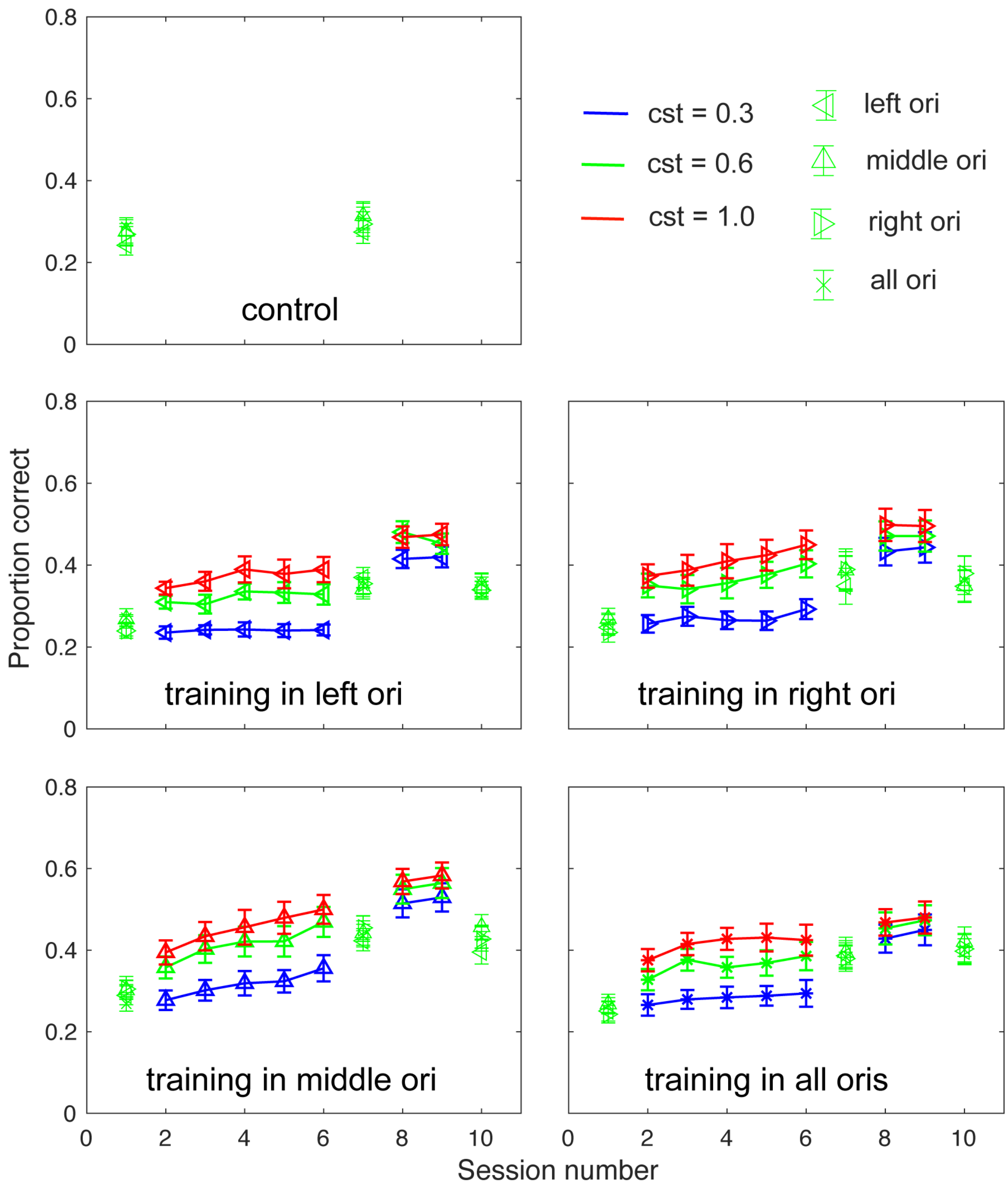

Figure 6. Behavioral results of the spatial-frequency identification task (Experiment 2). The five panels show results of the five groups: control (pre- and post-tests only, top left), training with a left orientation (middle left), a right orientation (middle right), a middle orientation (bottom left), and all orientations (bottom right). All pre- and post-tests were done with all five tested orientations displayed at the middle contrast (0.6). Training was done with three contrast levels (0.3, 0.6, and 1.0) in high external noise (sessions 2–6) or zero external noise (sessions 8 and 9). The control group changed little from pre- to post-test, whereas overall post-test performance of all orientations improved after training regardless of which orientations participants were trained on, signifying both learning and transfer (over a task-irrelevant feature).

again, no difference was observed between the two post-tests (most $p > 0.1$; see Supplementary Table S2-1 for details), so performance was combined for analysis. Analyses of variance found neither significant differences in pre-test performance among groups before training, $F(3, 36) = 1.040$, $p > 0.3$, nor differences in learning during training (session 2 to 6), $F(3, 36) = 1.608$, $p > 0.2$. A detailed analysis of the training sessions appears in Supplementary Table S2-2.

Taken together, all training groups started with similar initial performance and showed similar improvement over the five sessions of training. The pre-test to post-test improvement was also equivalent

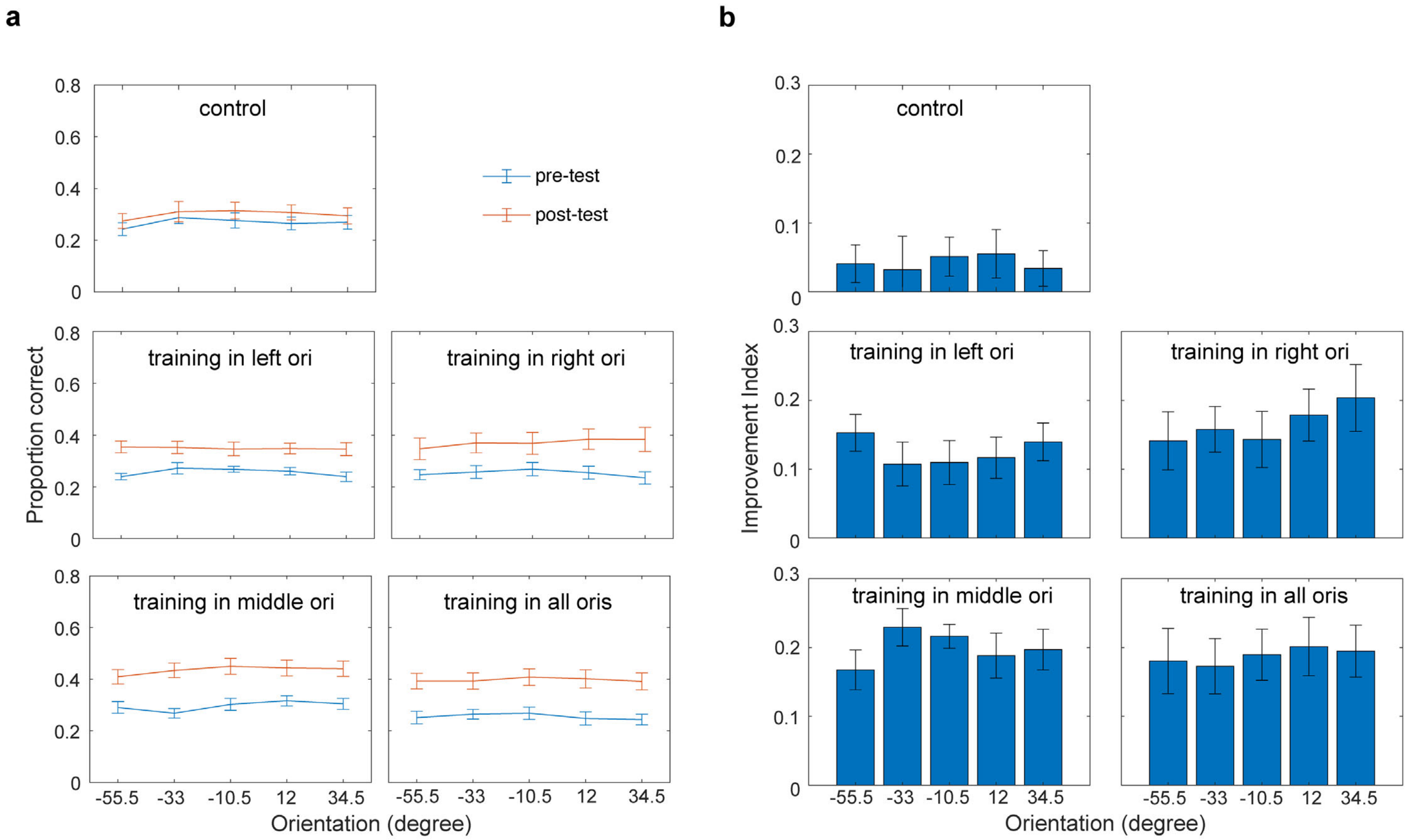

Figure 7. The pre- and post-tests accuracies and improvement indices of the spatial-frequency identification task (Experiment 2). Performance from the two post-tests were averaged. (**a**) The proportion correct results. The control group showed little improvement. In training groups, the (nearly) parallel line of the pre- and post-tests represents essentially complete transfer. Top left shows control; middle left, training in the left orientation; middle right, training in the right orientation; bottom left, training in the middle orientation; and bottom right, training in all five orientations. (**b**) Improvement indices comparing post- to pre-test performance (as defined in the text). Similar bar length across different orientations also signifies broad transfer over orientation of training. The error bars are *SEM*.

among the different training groups (see Supplementary Table S2-3). As before, all test orientations showed significant improvement from pre- to post-tests in all training groups (improvement greater than 0; all $p < 0.01$; see Supplementary Table S2-4 for details). In contrast, for the control group, no improvement was observed (all $p > 0.1$; see Figure 7b, top-left panel).

As in Experiment 1, we performed a Bayesian analysis that provided estimates of $p(H_0/D)$ and $p(H_1/D)$, or the probability of the null (or alternative) hypothesis given the data. Traditional p values reflect $p(H_0/D)$, or the probability of observing current data given the null hypothesis. The estimated effect sizes for group, stimulus orientation, and interaction were 0.244, 0.018, and 0.111, respectively, and the $p_{BIC}(H_0/D)$ values were 0.00, 1.00, and 1.00, respectively, reflecting a group difference between the control group and all of the training groups. When excluding the control group, the effect sizes were 0.0768, 0.0394, and 0.161 for group, orientation, and the interaction, respectively, and the $p_{BIC}(H_0/D)$ values were 0.77, 1.00, and 1.00, respectively, favoring the null hypothesis that the performance improvement due to learning occurred equivalently at all different stimulus orientations in all four training groups. (Corresponding standard frequentist statistical analysis is provided in Supplementary Table S2-3.) Consistent with Experiment 1, both Bayesian statistical assessments and frequentist analyses supported the null hypothesis: that learning in any of the orientations completely transfers to other orientations in the spatial-frequency identification task. The patterns of response confusion (the frequency of responses given to each stimulus) in the pre-tests compared to the post-tests for all five groups provided a consistent visualization of these findings, as shown in Supplementary Figures S2-1 (heatmaps) and S2-2 (response profiles).

Summary of experimental results

The goal of the current behavioral experiments was to test whether visual perceptual learning of a primary task trained at specific values of a task-irrelevant feature transferred across variations in that task-irrelevant feature. This was examined in two parallel studies: Experiment 1 trained observers in eight-alternative orientation identification with spatial frequency as the task-irrelevant feature, and Experiment 2 trained observers in eight-alternative spatial-frequency identification with orientation as the task-irrelevant feature.

Both studies showed broad transfer over variations in the task-irrelevant feature, even when initial training used only a single value of that feature. Although there were some minor deviations associated with specific stimuli, the overall findings were quite robust, with Bayesian statistics strongly favoring null effects of the task-irrelevant training feature on post-training improvements across the four training groups; that is, $p_{BIC}(H_0/D)$ values were near 1. Another important finding was the equivalence of learning and transfer in the four training groups as measured by the improvement index comparing post-training to pre-training assessments. Here, too, the Bayesian probabilities affirmatively support no substantial differences due to training protocol; that is, $p_{BIC}(H_0/D)$ values were near 1 for the effect of training group. Changes in the pattern of response confusions also generally supported these conclusions.

Overall, then, both experiments demonstrated generalization of visual perceptual learning across task-irrelevant feature variation and relative equivalence of the effects of training in the different training groups. Both findings intuitively support ideas of pooling over features and resulting feature-invariant representations even if only one example task-irrelevant feature was trained. The following section develops an extended version of the identification-integrated reweighting theory (the eI-IRT) to account for these behavioral effects of learning and transfer.

The extended I-IRT

Model description

The original integrated weighting theory (IRT) was first developed to account for generalization of perceptual learning over different retinal locations (Dosher et al., 2013). The IRT is a neural network model that makes perceptual decisions by weighting evidence (activation) in representation units connected to a decision unit and learns by changing those weights to improve performance—reweighting. The IRT uses augmented Hebbian learning rules to learn the reweighting with representation activations computed from a signal-processing front end inspired by early visual cortical processing, as in the augmented Hebbian reweighting model (AHRM) (Petrov et al., 2005; Petrov et al., 2006). The key insight of the IRT was to account for cross-location transfer with reweighting of location-invariant feature representations and specificity to location through location-specific feature representations.

We recently extended the reweighting models to account for multi-alternative (>2) identification in the I-IRT (Dosher et al., 2023; Liu et al., 2023) by implementing a decision structure with one mini-decision unit for each potential identification response, with the most active determining the response. This form of the reweighting theory accounted for the impact of different forms of feedback (no feedback, accuracy feedback, and response feedback) on learning in the eight-alternative orientation identification task (Liu et al., 2023) in both a fixed contrast design and a contrast threshold paradigm in spatial-frequency identification (using response feedback) (Dosher et al., 2023); both cases trained in a single value of the task-irrelevant dimension.

The I-IRT, like other reweighting models, has representation, decision, and learning modules (with feedback and criterion control units). The representation module has both location-specific and location-invariant representations for a given stimulus. The former is only active when a stimulus shows up at a specific retinal location, whereas the latter is active wherever the stimulus appears. Activations in both types of representations are weighted to each mini-decision unit (together with the activation in a criterion control unit). The identification response for each trial in the I-IRT is based on the most active mini-decision unit. During learning, feedback assists updating the weights between representations and the mini-decision units if external feedback is available. If there is no external feedback, the weights are updated according to the pure Hebbian rule (full response feedback was used here). The inclusion of location-invariant representations allows the model to generalize learning for stimulus in an untrained retinal location. For a full description of the original IRT, see Dosher et al. (2013), Liu et al. (2014), and Liu et al. (2015); for a full description of the I-IRT, see Dosher et al. (2023) and Liu et al. (2023).

Here, we further extended the I-IRT for broader transfer with feature-invariant representations (i.e., eI-IRT). The representations in the IRT and I-IRT are selective for both orientation and spatial frequency. For potential transfer over the task-irrelevant feature, we included new feature-invariant representations that pool over variations in the task-relevant feature (e.g., orientation) regardless of the level of another feature (e.g., spatial frequency). Activations in these feature-invariant representations, along with those

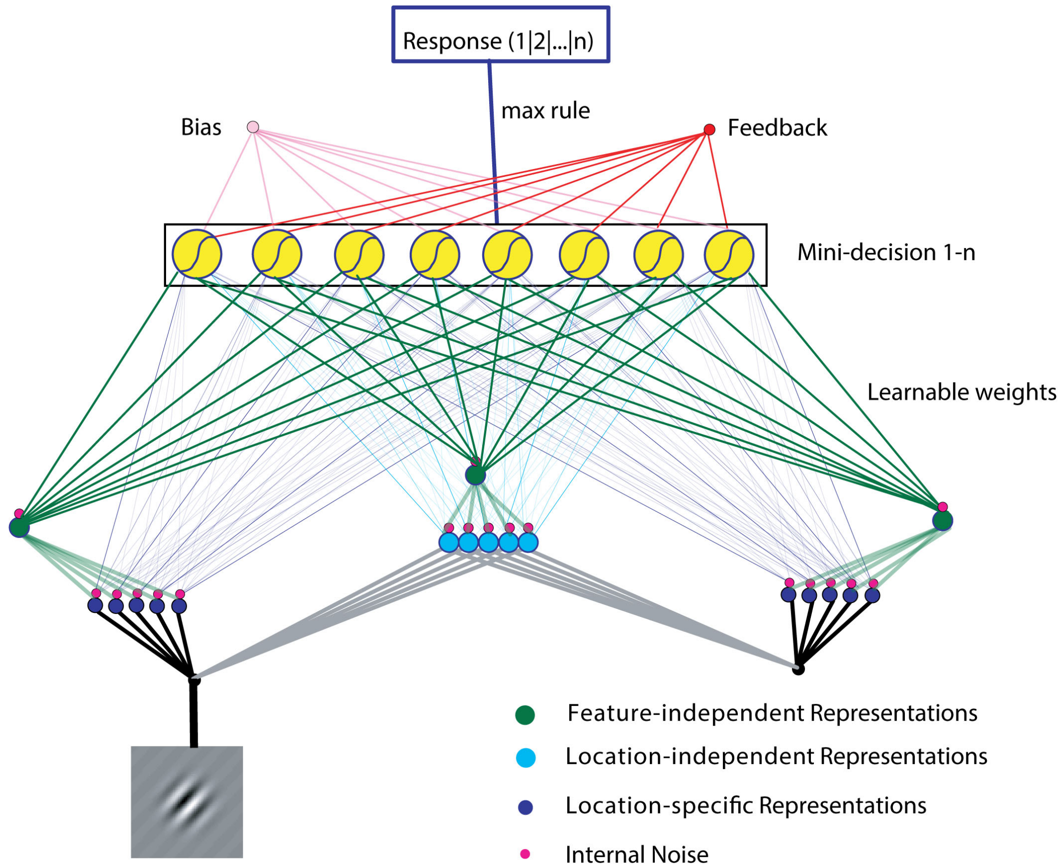

Figure 8. The eI-IRT model. The stimulus (here an oriented Gabor image) shown in one location is represented by location- and feature-specific representations, location-invariant representations, and feature-invariant representations. All of these representations are weighted to eight mini-decision units, one for each potential identification response. A max rule on the mini-decision activations determines the predicted response on each trial. Feedback is used to help updating the weights between representations and mini-decision units. Bias (or criterion control) is used to balance the response frequencies of different responses.

tuned to both spatial frequency and orientation, are fed into mini-decision units, and the weights are updated by learning. Through the action of the task-irrelevant feature-invariant representations, training an orientation task with one spatial frequency can potentially transfer to the same task with a different spatial frequency, and, similarly, a spatial frequency task can transfer over orientations. See Figure 8 for the model with feature-invariant representations. Versions of the IRT with and without the feature-invariant representations (units) were considered. A more detailed description of the model equations is provided in Supplementary Materials, Model Details.

Simulation fits to data

The extended I-IRT simulations were evaluated by running the simulation through the same experiments as the human observers: stimuli, trial sequence, and measurements. Predicted performance from the I-IRT (without feature-invariant representations) and the eI-IRT (with feature-invariant representations) was computed from the average of 100 simulated runs. Most parameters of the model were fixed a priori according to physiology or from previous implementations of the IRT and I-IRT (Dosher et al., 2013; Dosher et al., 2020; Dosher et al., 2023; Liu, Dosher, & Lu, 2014; Liu, Dosher, & Lu, 2015; Liu et al., 2023). Only the (single) model learning rate and internal noise terms (representational noise and decision noise) and a scaling factor of the stimulus activation were varied in the current quantitative fits to the behavioral data.

Crucially, in each experiment, model parameters were held the same across the four different training groups (low, middle, high, and all) and the control group, so that the difference in learning and transfer emerged organically from the model and experimental design, not from additional parameters for each training condition. The one exception was a scaling factor that varied slightly across different groups, set to fit slightly different initial performance in the first session due to random individual differences in subjects assigned to

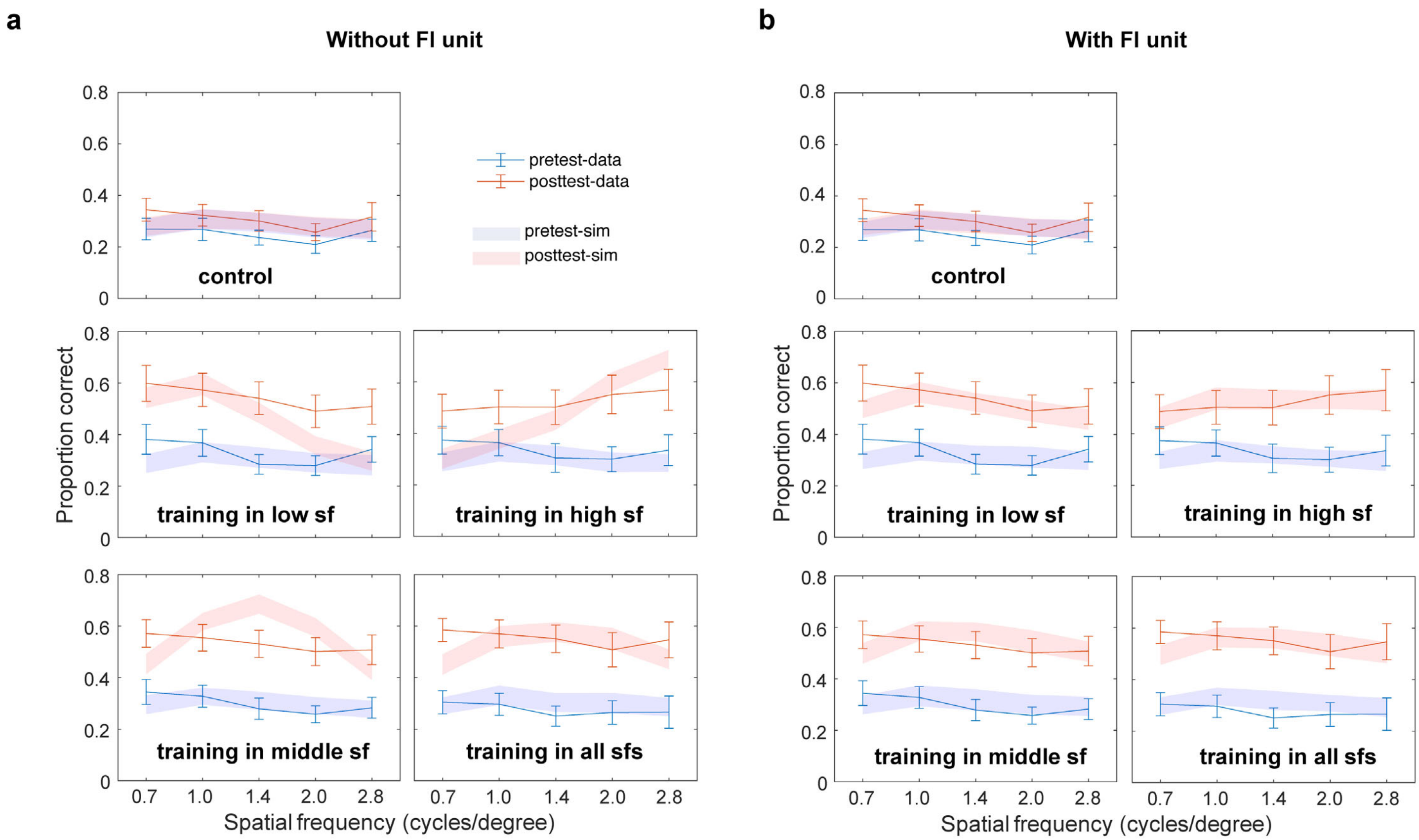

Figure 9. The I-IRT model simulation of the Experiment 1 data without FI units (**a**) and with FI units (**b**). As in Figure 5, the blue and red lines are the pre- and post-test data of the experiment, and the error bars are ± 1 *SEM*. Blue and red shaded areas are the pre- and post-test predictions of the simulation, and the shade height is ± 1 *SD*. (**a**) Without FI units, the I-IRT predicted specificity of the training not observed in the data, as training in low spatial frequency was predicted to transfer little to the high spatial frequency, and vice versa. (**b**) With FI units, the eI-IRT predicted wide generalization following training in a specific spatial frequency, in agreement with the data.

groups. As we will show, simulating the eI-IRT that incorporated feature-invariant representation units made predictions consistent with the behavioral results. In fitting the model simulations, we started with a grid search over the few variable parameters to find the likely regions for a good fit, then followed with an error minimization search starting with reasonable parameters in these regions. We searched for the parameters with a best least-square fit to the average data across all groups.

Fits of the I-IRT and eI-IRT to the behavioral results

We applied the I-IRT (Dosher et al., 2023; Liu et al., 2023), and the new eI-IRT to simulate the behavioral results. By simulating the behavioral experiments both with and without feature-invariant units, we sought to show that these units are both necessary and sufficient to account for our behavioral results. The addition of feature-invariant representation units pooling over the task-irrelevant dimension in the new eI-IRT parallels the use of location-invariant representations to account for location transfer in the original IRT (Dosher et al., 2013).

Figure 9a shows simulation predictions for pre-test and post-test data based on the I-IRT (Dosher et al., 2013; Liu et al., 2023) for the eight-alternative orientation identification task (Experiment 1). Varying values of noise terms, learning rate, and the scaling factor were selected to provide a good fit to the control group data, the initial session data in all training groups, and the average improvement from pre- to post-tests. The model included both location-dependent and location-invariant representation units tuned to combinations of orientation and spatial frequency content in the stimulus. Full task transfer over the irrelevant feature, however, was *not* predicted by the I-IRT model, which does not include the feature-invariant units. As shown in Figure 9a, without

Parameters	Parameter values			
Parameters set a priori				
Orientation spacing, $\Delta\theta$	15°			
Spatial frequency spacing, Δf	0.5 octave			
Maximum activation level, $A_{\max}$	1			
Weight bounds, $w_{\min}$ and $w_{\max}$	± 1			
Initial weights scaling factor, w_{init}	0.0338			
Activation function gain, γ	3.5			
Location-specific orientation bandwidth, $h\theta$	30°			
Location-invariant orientation bandwidth, $h_{\theta I}$	60°			
Location-specific frequency bandwidth, h_f	1 octave			
Location-invariant frequency bandwidth, h_{fI}	2 octave			
Radial kernel width, h_r	2 dva			
Parameters adjusted for the data	Without FI units		With FI units	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
Normalization constant k	1e-5			
Scaling factor a	0.036∼0.042	0.05∼0.06	0.018∼0.021	0.025∼0.032
Location-specific multiplicative noise, σ_m	0.1	0.05	0.1	0.05
Location-invariant multiplicative noise, $2^*\sigma_m$	0.2	0.1	0.2	0.1
Location-specific additive noise, σ_a	3e-6	0	3e-6	1e-6
Location-invariant additive noise, $2^*\sigma_a$	6e-6	0	6e-6	2e-6
Decision noise, σ_d	0.15	0.1	0.15	0.09
Learning rate, η	2.8e-4	2.5e-4	1.64e-4	
Bias weight, w_b	1.0			
Feedback weight, w_f	1.0			

Table 1. Parameters of the best-fitting model.

feature-invariant units the I-IRT consistently predicted more specificity to the task-irrelevant feature used during training than shown in the data; for example, training an orientation task in low spatial frequency is predicted to transfer only modestly to the high spatial frequency and vice versa.

We used the eI-IRT including feature-invariant (FI) units or representations to account for the broad transfer over variations in the irrelevant feature observed in our behavioral studies. Spatial frequency-invariant orientation representations pool the activations of all representations sensitive to a given orientation from all the spatial frequencies, and orientation-invariant spatial frequency representations pool over representations sensitive to a certain spatial frequency from all the orientations. See Figure 9b for the simulated predictions of this model with best-fit parameters, as listed in Table 1.

With spatial frequency-invariant orientation representations, the eI-IRT model simulations nicely captured the transfer results in Experiment 1 (compare Figure 9b, where $r^2 = 0.879$, with Figure 8a, where $r^2 = 0.627$), and, likewise with orientation-invariant spatial frequency representations, Experiment 2 results are also accounted for (compare Figure 10b, where $r^2 = 0.886$ with Figure 9a, where $r^2 = -0.311$). The model without FI unit predicted strong specificity and yielded a very low r^2. See Supplementary Materials, Model Details, for further details and model discussion.

A look at the weight dynamics, or changes in weights over training, shows how the model embodies both learning and transfer with or without the feature-invariant units (Figure 11). In these figures, weights from representations that are most relevant to the task (i.e., supporting the identification of a particular spatial frequency or orientation) are plotted, and an increase in these weights benefits the task. The left panels summarize weight changes for a model without the feature-invariant units (see Figures 9a and 10a for performance simulation results), and the right panels summarize weight changes for a model with the FI units (see Figures 9b and 10b for performance simulation results). The top panels are for Experiment 1 (orientation identification), and the bottom ones are for Experiment 2 (spatial frequency identification). For each panel, each bar shows post-training weights on selected representations to best reveal how weights changed to improve performance. These graphs show those most closely matched to the task-relevant feature for each of three values of the task-irrelevant value, with dashed lines indicating initial pre-training weights. For example, for orientation identification, Figure 11 shows the weight on representations tuned to the closest orientation for each stimulus (averaged over

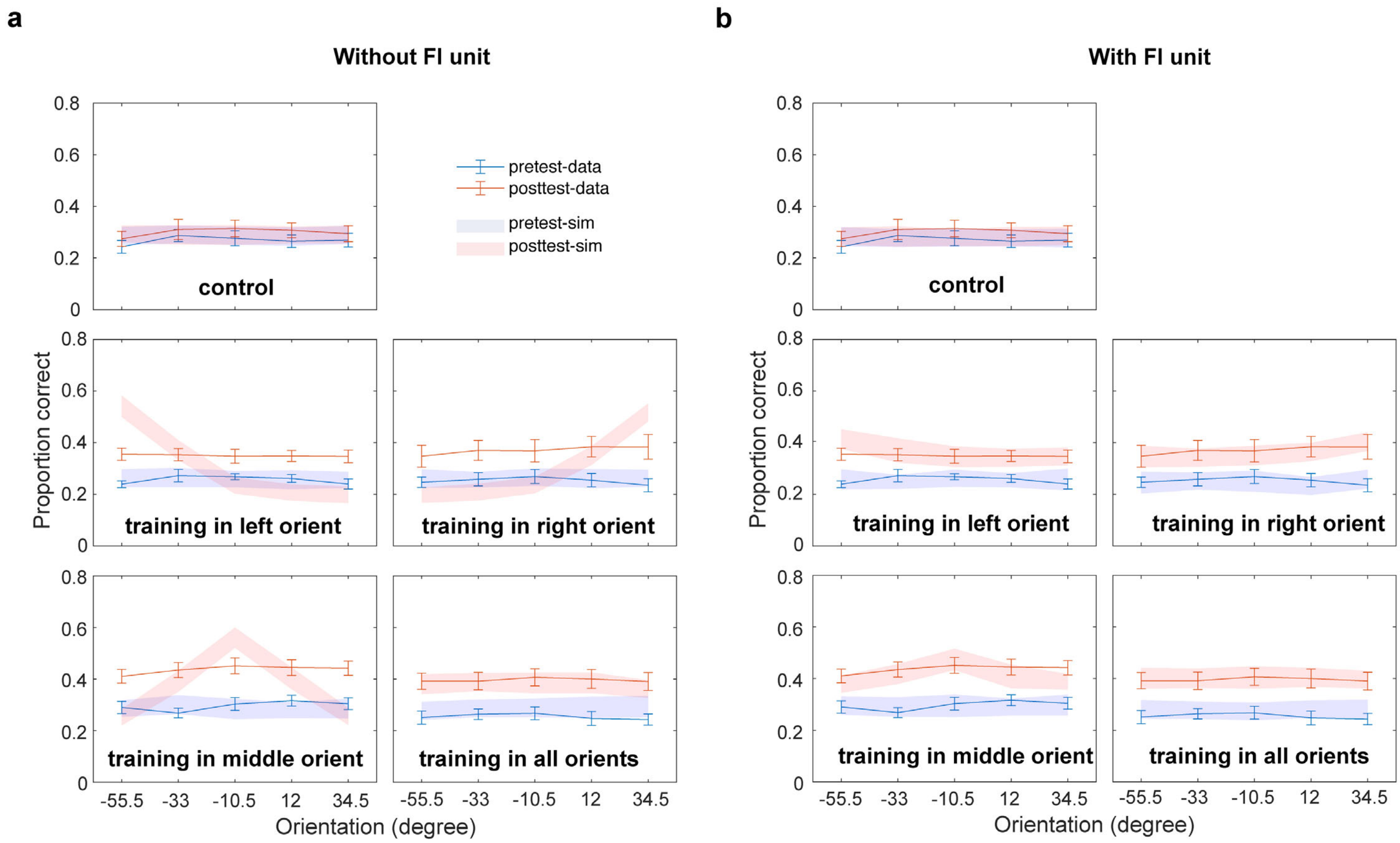

Figure 10. The I-IRT model simulation of the Experiment 2 data without FI units (**a**) and with FI units (**b**). Similar to Figure 7, the blue and red lines are the pre- and post-tests of the experiment and the error bars are ±1 *SEM*. Blue and shaded areas are the pre- and post-tests of the simulation, and the shade height is ±1 *SD*. (**a**) Without the FI units, the I-IRT overpredicted the specificity of the training, as training in a specific orientation (left, middle, or right) transferred little to orientations farther away. (**b**) With the FI units, the I-IRT predicted more generalization of training in a specific orientation, in agreement with the data.

stimuli) and also tuned to low, medium, and high spatial frequencies, and for the FI units that pool over them for the extended I-IRT model.

For Experiment 1 (Figures 11a and 11b), in all training groups, the weights from the standard feature-specific representations to the decision units increased in amplitude for the most relevant features (matching in both orientation and spatial frequency), which supports both learning in the training task and partial specificity when trained with one spatial frequency and tested on another far away from the trained one (i.e., trained with the highest spatial frequency and tested with the lowest spatial frequency and vice versa). Note that the activation profile for the low spatial frequency stimulus is broader with lower peak values than that for the high spatial frequency (consistent with Fourier analysis of the stimuli), so the post-training weights of low spatial-frequency channels are also of lower values and more spread than those of high spatial-frequency channels (see Figure 11a, comparing the blue bar in the middle left subpanel with the red bar in the middle right subpanel). When the FI unit is available as in Figure 11b, the weights from the most relevant orientation-specific but feature-invariant (spatial-frequency invariant) representations to decision units increase in amplitude regardless of which spatial frequency is trained. It is these increased weights on FI representations that drive transfer to the stimuli with non-trained spatial frequencies. Similar patterns hold for the simulation for the spatial frequency task. Improvement in task performance came from changes in weights for the standard feature-specific representations (Figures 11c and 11d) and also from changes in weights for the orientation-invariant spatial-frequency representations (Figure 11d) when available, which is of benefit to tests for stimuli for trained and untrained orientations. Therefore, the addition of feature-invariant representations is both necessary and sufficient to account for the broad transfer of learning effects over a task-irrelevant feature in the context of the eI-IRT model.

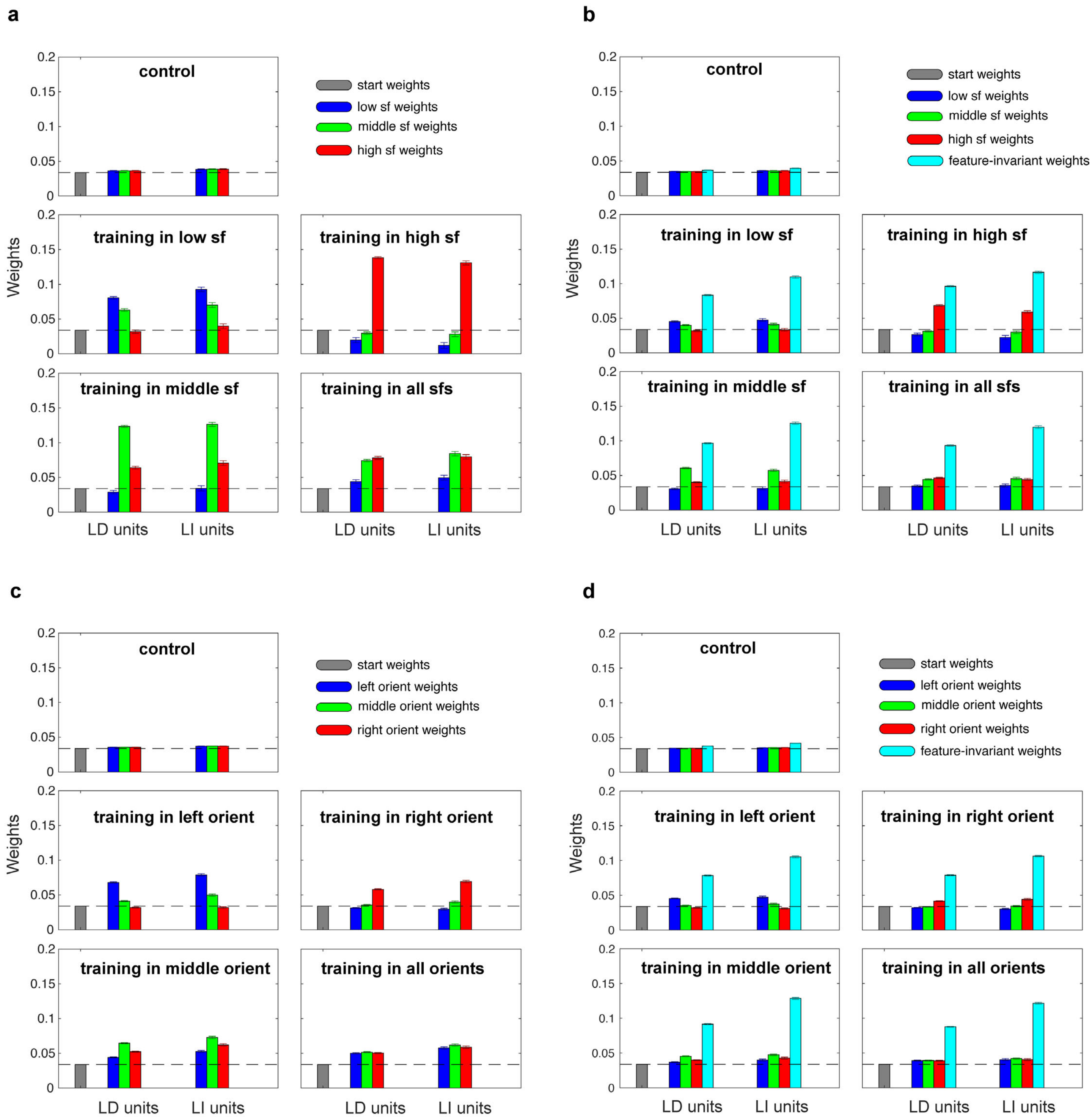

Figure 11. For Experiment 1, the weight dynamics of the I-IRT without (**a**) and with (**b**) FI units are shown; for Experiment 2, the most relevant representation channels comparing post-training weights to pre-training weights (**c**, **d**) are shown. (**a**) Pre- and post-training weights for the fit to the orientation identification task without FI units; when training at a specific spatial frequency, the weights for channels most selective to that spatial frequency increased the most, and weights from other channels increased less or even decreased. This change of feature-specific weights enables learning of the practiced spatial frequency but limits transfer to other spatial frequencies. (**b**) Pre- and post-training weights in the fit with FI units. Besides similar weights change pattern as in (a) from feature-specific units, the weights for feature-invariant channel increased regardless of which spatial frequency was trained. The increase of both feature-specific and feature-invariant weights enables learning at the trained spatial frequency, whereas an increase of the FI weights enables transfer across spatial frequencies. This pattern is the same for both location-specific and location-invariant

→

←
units. (**c**, **d**) The equivalent weights change for the spatial-frequency identification task. Similarly, the increased weights of both feature-specific and feature-invariant weights enable learning, whereas the weights increase of the feature-invariant unit enables transfer across different orientations.

Discussion

Transfer over task-irrelevant feature variation in the stimulus

In this project, we proposed and tested a general mechanism by which transfer across task-irrelevant features might occur in visual perceptual learning. Inspired by visual psychophysical studies of how the components of compound stimuli affect pattern discrimination (e.g., Olzak & Thomas, 1999) and by physiology, as well as modeling approaches, we proposed that a "pooling" mechanism drives feature-invariant representations that are the basis of learning and transfer over task-irrelevant features. We tested this idea by studying perceptual learning in eight-alternative identification tasks in two parallel experiments: orientation identification and spatial-frequency identification. Both experiments resulted in almost complete transfer across variations in the task-irrelevant dimension regardless of the stimulus (I) used during training. A recent paper also reported transfer of learned orientation discrimination across task-irrelevant variation in spatial frequency in a two-interval discrimination (with standard) task (Manenti et al., 2023).

Task-relevant transfer compared to task-irrelevant transfer

The literature on transfer over task-relevant feature provides a context for transfer over variations in task-irrelevant stimulus features. First, Dosher et al. (2013) found that training in two-alternative orientation identification transferred more over a task-irrelevant feature (location) than over the relevant one (orientation)—in which, because the original and transfer orientations were sufficiently different, any transfer likely reflects general learning to exclude external noise as well other possible details. Second, variation in task-relevant features during training leads to significant disruption of learning in two-alternative tasks (Dosher et al., 2020; Parkosadze, Otto, Malania, Kezeli, & Herzog, 2008; Sagi, Adini, Tsodyks, & Technion, 2003; Zhang et al., 2008), sometimes referred to as the roving deficit. For example, training orientation discrimination at multiple reference orientations was found to substantially reduce perceptual learning (Dosher et al., 2020), and training contrast discrimination at multiple reference contrasts showed no perceptual learning (Kuai, Zhang, Klein, Levi, & Yu, 2005). However, when both irrelevant feature (location) and relevant feature (reference orientation) were varied in double training experiments, a two-interval orientation discrimination task showed substantial transfer (Xie & Yu, 2020). Taken together, these observations suggest that phenomena observed for task-relevant stimulus variation may not apply in the same way (or at all) for variations in task-irrelevant dimensions.

Integrated reweighting theory of learning and transfer

A computational model, the extended identification integrated reweighting theory (eI-IRT), developed here, showed that feature-invariant representations in the stimulus encoding, analogous to the location-invariant representations that previously accounted for transfer over retinal locations (Dosher et al., 2013; Dosher et al., 2020), were necessary and sufficient to account for such results. Learned weights from these invariant representation units to the decision units predict broad transfer over a task-irrelevant feature, based on reweighting of connections from the task-irrelevant feature-invariant representations. At the same time, the original IRT makes systematic predictions about transfer over switches in task-relevant task stimuli, as well as over variations in a task-irrelevant dimension (Dosher et al., 2013). It also predicts the reduction or elimination of learning in designs that rove the task-relevant stimuli (Dosher et al., 2020), and related models account for many results from double training paradigms (Taluri et al., 2015).

Pooling mechanisms in psychophysical models

Pooling mechanisms such as the feature-invariant representations in the eI-IRT may be related to pooling phenomena previously suggested by a set of psychophysical experiments from Olzak and colleagues ("cigars" for spatial frequency-pooled orientation representations and "doughnuts" for orientation-pooled spatial frequency representations; see Olzak & Thomas, 1992; Olzak & Thomas, 1999; Olzak & Wickens, 1997; Thomas & Olzak, 1996). Although their experiments focused on pattern discrimination of compound stimuli after some practice and the different orientations or spatial frequencies leading to pooling were simultaneously present in their compound stimuli, our experiments examined perceptual learning given a single training stimulus on each trial and the generalization of learning over the task-irrelevant feature in memory. The eI-IRT models this via learned weights for feature-invariant representations. A common integration/pooling mechanism may account for our results and those of Olzak and colleagues. Although there may be differences in the detailed mechanisms in the two sets of studies, a pooling process seems to underlie both. The availability of feature-invariant representations enables focus on important features for the overt task while also supporting potential transfer across stimulus variations.

Physiological substrates

Macaque monkey V1 and V2 neurons have a diverse selectivity to both orientation and spatial frequency (De Valois, Albrecht, & Thorell, 1982), which may be adaptive for processing natural images: narrowly tuned orientation neurons and broadly tuned neurons for stimuli with multiple orientations (Goris, Simoncelli, & Movshon, 2015). Many neurons are selective for both orientation and spatial frequency in primary visual cortex in macaque, yet neurons that are narrowly tuned to orientation but broadly tuned to spatial frequency, and vice versa, also may exist (DeValois et al., 1982). Also, similar tuning of neurons to orientation and spatial frequency seems to be observed in V2 and V4 (Zhang, Schriver, Hu, & Roe, 2023). Other models impute a circle/cylinder arrangement for orientation and spatial frequency coding in primary visual cortex (Bressloff & Cowan, 2003), analogous to the doughnut/cigar pooling model, but do not fully specify the physiological underpinnings of any pooling or summation.

Olzak and Thompson (1999, p. 251) said, “The neural substrates of the summing [e.g., pooling] circuits we describe are entirely unknown, although it is tempting to draw parallels between the psychophysics and physiological findings.” They point to summation within orientation columns, which include cells tuned to different spatial frequencies, as a basis for the “cigars” in their model, arguing that these “may be represented in V1 itself, or … in V2 or beyond.” The substrate underlying summations of spatial frequency over orientations is less obvious, but they suggest that long cortical slabs of activation in experiments testing a single spatial frequency across orientations might support the “doughnuts” in their model (Olzak & Thompson, 1999). Absent significant new physiological evidence, we concur that selective tuning for the task-relevant feature across multiple values of the task-irrelevant feature at an early visual cortical level cannot be ruled out.

Alternatively, it may be that evidence across task-irrelevant variation is pooled at a higher level of cortex through readout mechanisms. It has been argued, for example, that invariant representations for orientation pooled over spatial frequency arise, especially with varied training, and are localized in ventral temporal cortex (VTC) (Manenti et al., 2023). Notably, it has also been argued that the judgment process and pooling might differ for coarse or fine discriminations (Adab & Vogels, 2016). The location-invariant representations of the IRT, for example, have been associated with V4 or higher levels of visual cortex (Dosher et al., 2013; Talluri et al., 2015).

To summarize, invariant (pooled) representations for orientation (over spatial frequency) and spatial frequency (over orientation) may (or may not) occur at the same early cortical level (V1/V2) as the basic units of the IRT that are tuned jointly for spatial frequency and orientation, whereas the location-invariant units have been associated with higher levels of the visual hierarchy (V4/IT) (Talluri et al., 2015). We believe that identification of objects via complex feature arrays surely will reflect decision combinations in higher brain regions (Dosher & Lu, 2020). Further understanding of the organization of selectivity in visual cortex from cellular physiology or brain imaging may be required to fully understand physiological substrates for task-irrelevant feature invariant learning.

Different factors affecting learning and transfer

The degree of transfer across either task-irrelevant or task-relevant stimulus variations may be influenced by the details of the testing paradigms, such as the variability in stimuli, number of alternatives, number of presentation intervals, or granularity of discrimination (coarse or fine). For example, a prior review (Raviv et al., 2022) argued that variability in both stimuli and training schedule shapes learning and generalization. Specifically, the authors suggested that variability in an irrelevant dimension early in training may help refine decision boundaries and boost later learning and generalization. Broadly similar claims have been made for double training experiments (e.g., Xiao et al., 2008; Zhang et al., 2010), in which training in multiple spatial locations and tasks has been reported to increase generalization of learning across locations (a task-irrelevant dimension) and tasks (a task-relevant dimension). In contrast, classical visual perceptual learning experiments that trained in one retinal location or feature often showed substantial specificity (e.g., Karni & Sagi, 1991).

The study of perceptual learning of orientation (two-interval, standard first two-alternative discrimination thresholds) cited above (Manenti et al., 2023) reported that training with variable spatial frequencies in a coarse discrimination task showed more transfer than those trained using a single spatial frequency. The authors also argued that feature variations during training similarly leads to more generalization in a deep learning network. In contrast, Experiments 1 and 2 in the current study, which tested eight-alternative identification, showed generalization without varying the task-irrelevant feature during training; three groups trained with single (different) values of a task-irrelevant feature showed approximately equivalent performance and the same wide transfer as the group trained with a mixture of values of the task-irrelevant dimension.

One potential influence is the use of a pre-test baseline that included variations in the task-irrelevant feature in our experiments, as one review suggested that the use of pre-test baselines increased transfer compared to designs that did not use pre-tests (Dosher & Lu, 2020, chapter 3). However, a study design similar to that of our experiments (pre- and post-tests at multiple spatial frequencies but training at one) in two-interval contrast detection found narrower spatial frequency tuning following training in normal control observers (Huang et al., 2008; Zhou et al., 2015).

Another possibility is that the two-alternative and *n*-alternative tasks may yield somewhat different sensitivity to factors such as variation in trained stimuli. The eight-alternative identification tasks used here require either eight templates (as in the I-IRT) or seven criteria, whereas the two-alternative identification tasks require either two templates or one criterion in the decision system. Some might argue that the two tasks have different memory load. In our view, however, the two-alternative and *n*-alternative identification designs have the same memory demands during an individual trial, where the observer sees and responds to only a single test stimulus on each trial, and it is the decision structure that is more complex. There is no delay imposed between the stimulus presentation and the response (unlike visual memory tasks that manipulate a delay after the stimulus display). On the other hand, the typical two-interval designs present a standard and another stimulus in two different intervals with a brief interval between, followed by a response either immediately (e.g., Huang et al., 2008) or after a significant delay (Manenti et al., 2023). This requires other forms of memory during the trial. (The two-interval designs require specific modeling of the memory delay and the kinds of stimulus information retained; see Liu, Lu, & Dosher, 2021.) Additionally, *n*-alternative identification designs tend to use more coarsely varying stimuli, whereas two-interval designs likely measure minimal stimulus difference thresholds and so may require reliance on all features of the stimulus to achieve good performance.

Ideally, each detailed experimental design could be modeled within a computational framework such as the IRT or other model framework to promote a full understanding of the influence of these various factors. Depending on practical goals (whether specificity or generalization is desired), we may be able to use these models to design different experimental manipulations with knowledge from all these learning and transfer studies.

Conclusions

In summary, we showed broad generalization of learning across task-irrelevant features in visual perceptual learning and propose a model to account for this transfer. The proposed distinction between transfer over task-relevant features versus task-irrelevant features may be a way to both organize and understand sometimes apparently conflicting results on transfer in perceptual learning. It may also provide insights relevant to the design of effective training paradigms. Further research in transfer both across irrelevant and relevant features will shed light on brain mechanisms of learning and instruct practical training paradigms.

Keywords: perceptual learning, transfer, generalization, task-irrelevant feature

Acknowledgments

Supported by a grant from the National Eye Institute, EY-17491.

Commercial relationships: none.
Corresponding author: Barbara Dosher.
Email: bdosher@uci.edu.
Address: Department of Cognitive Sciences, University of California, Irvine, CA 92617, USA.

References

Adab, H. Z., & Vogels, R. (2016). Perturbation of posterior inferior temporal cortical activity impairs coarse orientation discrimination. *Cerebral Cortex, 26*(9), 3814–3827.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*(6), 953–965.

Bressloff, P. C., & Cowan, J. D. (2003). A spherical model for orientation and spatial–frequency tuning in a cortical hypercolumn. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1438), 1643–1667.

Connolly, K. (2019). *Perceptual learning: The flexibility of the senses*. Oxford, UK: Oxford University Press.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology, 78*(6), 2889–2894.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*(5), 545–559.

Dosher, B., Jeter, P., Liu, J., & Lu, Z.-L. (2013). An integrated reweighting theory of perceptual learning. *Proceedings of the National Academy of Sciences, USA, 110*(33), 13678–13683.

Dosher, B., Liu, J., Chu, W., & Lu, Z.-L. (2020). Roving: The causes of interference and re-enabled learning in multi-task visual raining. *Journal of Vision, 20*(6):9, 1–26, https://doi.org/10.1167/jov.20.6.9.

Dosher, B., Liu, J., & Lu, Z.-L. (2023). Learning spatial frequency identification through reweighted decoding. *Journal of Vision, 23*(6):3, 1–31, https://doi.org/10.1167/jov.23.6.3.

Dosher, B. A., & Lu, Z.-L. (2005). Perceptual learning in clear displays optimizes perceptual expertise: Learning the limiting process. *Proceedings of the National Academy of Sciences, USA, 102*(14), 5286–5290.

Dosher, B., & Lu, Z.-L. (2007). The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychological Science, 18*(6), 531–539.

Dosher, B., & Lu, Z.-L. (2020). *Perceptual learning: How experience shapes visual perception*. Cambridge, MA: MIT Press.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision, 2*(2):5, 190–203, https://doi.org/10.1167/2.2.5.

Goris, R. L., Simoncelli, E. P., & Movshon, J. A. (2015). Origin and function of tuning diversity in macaque visual cortex. *Neuron, 88*(4), 819–831.

Green, C. S., Banai, K., Lu, Z.-L., & Bavelier, D. (2018). Perceptual learning. In J. T. Wixted, E. A. Phelps & L. Davachi (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, learning and memory* (Vol. *2*, pp. 1–47). New York: John Wiley & Sons.

Green, C. S., Kattner, F., Siegel, M. H., Kersten, D., & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *Journal of Vision, 15*(10):5, 1–14, https://doi.org/10.1167/15.10.5.

Hou, F., Huang, C. B., Tao, L., Feng, L., Zhou, Y., & Lu, Z.-L. (2011). Training in contrast detection improves motion perception of sinewave gratings in amblyopia. *Investigative Ophthalmology & Visual Science, 52*(9), 6501–6510.

Hou, F., Lesmes, L., Bex, P., Dorr, M., & Lu, Z.-L. (2015). Using 10AFC to further improve the efficiency of the quick CSF method. *Journal of Vision, 15*(9):2, 1–18, https://doi.org/10.1167/15.9.2.

Hu, D. Z., Wen, K., Chen, L. H., & Yu, C. (2021). Perceptual learning evidence for supramodal representation of stimulus orientation at a conceptual level. *Vision Research, 187*, 120–128.

Huang, C. B., Zhou, Y., & Lu, Z.-L. (2008). Broad bandwidth of perceptual learning in the visual system of adults with anisometropic amblyopia. *Proceedings of the National Academy of Sciences, USA, 105*(10), 4068–4073.

Jeter, P. E., Dosher, B., Liu, S. H., & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research, 50*(19), 1928–1940.

Jeter, P. E., Dosher, B., Petrov, A., & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision, 9*(3):1, 1–13, https://doi.org/10.1167/9.3.1.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences, USA, 88*, 4966–4970.

Kuai, S. G., Zhang, J. Y., Klein, S. A., Levi, D. M., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience, 8*(11), 1497–1499.

Liu, J., Dosher, B., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research, 99*, 46–56.

Liu, J., Dosher, B., & Lu, Z.-L. (2015). Augmented Hebbian reweighting accounts for accuracy and induced bias in perceptual learning with reverse feedback. *Journal of Vision, 15*(10):10, 1–21, https://doi.org/10.1167/15.10.10.

Liu, J., Lu, Z.-L., & Dosher, B. (2021). Multi-location, two-interval paradigms can overcome roving costs–an explanation of Xie & Yu (2020) data by an extended Integrating Reweighting Theory (IRT). *Journal of Vision, 21*(9), 2264, https://doi.org/10.1167/jov.21.9.2264.

Liu, J., Lu, Z.-L., & Dosher, B. (2023). Informational feedback accelerates learning in multi-alternative perceptual judgements of orientation. *Vision Research, 213*, 108318.

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research, 40*(1), 97–109.

Lu, Z.-L., & Dosher, B. (2013). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: MIT Press.

Lu, Z.-L., & Dosher, B. A. (2022). Current directions in visual perceptual learning. *Nature Reviews Psychology, 1*(11), 654–668.

Manenti, G. L., Dizaji, A. S., & Schwiedrzik, C. M. (2023). Variability in training unlocks generalization in visual perceptual learning through invariant representations. *Current Biology, 33*(5), 817–826.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679–690.

Olzak, L. A., & Thomas, J. P. (1992). Configural effects constrain Fourier models of pattern discrimination. *Vision Research, 32*(10), 1885–1898.

Olzak, L. A., & Thomas, J. P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research, 39*(2), 231–256.

Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: Orientation information is integrated across spatial scale; spatial-frequency and contrast information are not. *Perception, 26*(9), 1101–1120.

Parkosadze, K., Otto, T. U., Malania, M., Kezeli, A., & Herzog, M. H. (2008). Perceptual learning of bisection stimuli under roving: Slow and largely specific. *Journal of Vision, 8*(1):5, 1–8, https://doi.org/10.1167/8.1.5.

Petrov, A. A., Dosher, B., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review, 112*(4), 715–743.

Petrov, A. A., Dosher, B., & Lu, Z.-L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research, 46*(19), 3177–3197.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences, 26*(6), 462–483.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research, 51*(13), 1552–1566.

Sagi, D., Adini, Y., Tsodyks, M., & Technion, A. W. (2003). Context dependent learning in contrast discrimination: Effects of contrast uncertainty. *Journal of Vision, 3*(9), 173, https://doi.org/10.1167/3.9.173.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *The Journal of Physiology, 483*(3), 797–810.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics, 52*(5), 582–588.

Sowden, P. T., Rose, D., & Davies, I. R. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*(10), 1249–1258.

Talluri, B. C., Hung, S. C., Seitz, A. R., & Series, P. (2015). Confidence-based integrated reweighting model of task-difficulty explains location-based specificity in perceptual learning. *Journal of Vision, 15*(10):17, 1–12, https://doi.org/10.1167/15.10.17.

Thomas, J. P., & Olzak, L. A. (1996). Uncertainty experiments support the roles of second-order mechanisms in spatial frequency and orientation discriminations. *Journal of the Optical Society of America A*, *13*(4), 689–696.

Vogels, R., & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision Research, 25*(11), 1679–1687.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804.

Wang, R., Wang, J., Zhang, J. Y., Xie, X. Y., Yang, Y. X., Luo, S. H., . . . Li, W. (2016). Perceptual learning at a conceptual level. *Journal of Neuroscience, 36*(7), 2238–2246.

Watanabe, T., Náñez Sr, J. E., Koyama, S., Mukai, I., Liederman, J., & Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nature Neuroscience, 5*(10), 1003–1009.

Xiao, L. Q., Zhang, J. Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18*(24), 1922–1926.

Xie, X. Y., & Yu, C. (2020). A new format of perceptual learning based on evidence abstraction from multiple stimuli. *Journal of Vision, 20*(2):5, 1–9, https://doil.org/10.1167/jov.20.2.5.

Xiong, Y. Z., Zhang, J. Y., & Yu, C. (2016). Bottom-up and top-down influences at untrained conditions determine perceptual learning specificity and transfer. *eLife, 5*, e14614.

Zhang, J. Y., Kuai, S. G., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2008). Stimulus coding rules for perceptual learning. *PLoS Biology, 6*(8), e197.

Zhang, Y., Schriver, K. E., Hu, J. M., & Roe, A. W. (2023). Spatial frequency representation in V2 and V4 of macaque monkey. *eLife, 12*, e81794.

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

Zhou, J., Yan, F., Lu, Z.-L., Zhou, Y., Xi, J., & Huang, C. B. (2015). Broad bandwidth of perceptual learning in second-order contrast modulation detection. *Journal of Vision, 15*(2):20, 1–10, https://doi.org/10.1167/15.2.20.

Zhou, Y., Huang, C., Xu, P., Tao, L., Qiu, Z., Li, X., . . . Lu, Z.-L. (2006). Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. *Vision Research, 46*(5), 739–750.