Specificity of perceptual learning increases with increased training

Pamela E. Jeter [a,d], Barbara Anne Dosher [a,*], Shiau-Hua Liu [a,b], Zhong-Lin Lu [c]

[a] Memory, Attention and Perception Laboratory (MAP), Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA
[b] Counseling and Clinical Psychology Department, National Dong Hwa University, Hualien 97401, Taiwan
[c] Laboratory of Brain Processes (LOBES), Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA 90089-1061, USA
[d] Lions Vision Research and Rehabilitation Center, Department of Ophthalmology, Johns Hopkins University, Baltimore, MD 21205, United States

ABSTRACT

Perceptual learning often shows substantial and long-lasting changes in the ability to classify relevant perceptual stimuli due to practice. Specificity to trained stimuli and tasks is a key characteristic of visual perceptual learning, but little is known about whether specificity depends upon the extent of initial training. Using an orientation discrimination task, we demonstrate that specificity follows after extensive training, while the earliest stages of perceptual learning exhibit substantial transfer to a new location and an opposite orientation. Brief training shows the best performance at the point of transfer. These results for orientation–location transfer have both theoretical and practical implications for understanding perceptual expertise.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Visual perceptual learning refers to improvements that develop expertise, through practice, in distinguishing differences in visual sensory features, such as contrast (Yu, Klein, & Levi, 2004), orientation (Dosher & Lu, 1998, 1999; Jeter, Dosher, Lu, & Petrov, 2009) or position (Ahissar & Hochstein, 1997; Karni & Sagi, 1991), see Fahle and Poggio (2002), for a review). Perceptual learning differs in magnitude depending upon the task (Fine & Jacobs, 2002) and may be most useful as a training or therapeutic tool if it generalizes or *transfers* to other similar visual tasks or attributes, such as spatial position, orientation and spatial frequency (Huang, Zhou, & Lu, 2008; Li, Polat, Makous, & Bavelier, 2009; Polat, 2010). More often, however, *specificity* (failure of transfer) to trained stimulus attributes or tasks is cited, although *partial* specificity and *partial* transfer (Ahissar & Hochstein, 1997; Ramachandran & Braddick, 1973) are often observed. The specificity of visual perceptual learning is the trademark finding that has led many researchers to infer that experience-dependent training alters representations in early visual cortex in areas with small receptive fields that are selective for orientation and position (Fahle & Poggio, 2002; Karni & Sagi, 1991; Gilbert, Sigman, & Crist, 2001).

What determines the extent of (partial) transfer or specificity? Knowing how specificity develops may greatly improve our understanding of perceptual learning. Specificity has typically been assessed only after the initial training approaches asymptotic level.

Here we ask, does specificity develop over training, and if so, how? We hypothesized that the amount of training on the initial task can be a critical factor determining the specificity of perceptual learning. To our knowledge, there are no prior studies that have manipulated the amount of initial training and measured subsequent specificity of visual perceptual learning. In this study, we manipulate the number of blocks of training over sessions (and not the number of trials in block, i.e., Censor & Sagi, 2009). Using a high precision orientation discrimination task, and transfer of learning to a new orientation and location, we show that specificity is a dynamic property, and that the extent and nature of specificity depends critically upon the extent of initial training.

1.1. Specificity and transfer

Early reports of the extraordinary specificity of visual perceptual learning to the trained stimuli, for example in a texture search task (Karni & Sagi, 1991), led to the conclusion that learning in different locations or for different stimuli occurs in independent representations, perhaps corresponding with V1 cells that code relatively precisely for retinal position and orientation (Fiorentini & Berardi, 1981; Gilbert et al., 2001). One classic study of orientation discrimination (Schoups, Vogels, & Orban, 1995) found specificity to different retinal positions, following some transfer from an initial training at fovea. Other studies (Crist, Kapadia, Westheimer, & Gilbert, 1997; Shiu & Pashler, 1992) found specificity to trained orientations. A number of these cases showed partial specificity and partial transfer (Beard, Levi, & Reich, 1995; Fahle & Poggio, 2002). Other studies of perceptual learning in visual search, motion
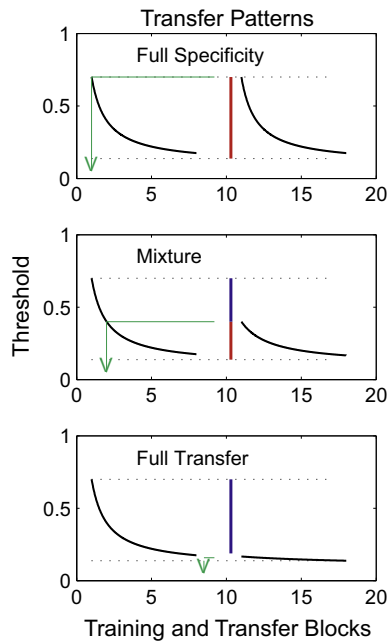
Fig. 1. A schematic illustration shows full specificity (top), partial transfer/partial specificity (middle) or full transfer (bottom). Black lines show hypothetical improvements in a threshold performance measure with perceptual learning, with the learning curves on the left for initial training, and the learning curves on the right for transfer phase training. Blue vertical lines mark the improvements in performance due to transfer, while red vertical lines mark the converse failure to transfer, or specificity. The transfer can also be characterized in terms of the equivalent amount of practice required to yield the performance level at the point of task switch, shown by the green lines dropped to the practice axis at the equivalent blocks of learning at the point of transfer.

direction discrimination, and orientation judgments (Ahissar & Hochstein, 1997; Jeter et al., 2009; Liu & Weinshall, 2000) found that specificity was a property of more demanding, high precision tasks (i.e., discrimination between very similar orientations), while transfer occurs more readily for low precision tasks. Recent studies also suggest that transfer may be greater if the tasks retain common judgment properties (Webb, Roach, & McGraw, 2007), or if a location has been previously trained in another task (Xiao et al., 2008; Zhang, Xiao, Klein, Levi, & Yu, 2010).

Transfer is the improvement in performance – here contrast threshold – in a new task at the point of the task switch, relative to an untrained baseline[1] (see Dill (2002), for a general discussion). Full transfer occurs when performance after the task switch continues at the trained level of the initial task. Specificity is the return towards untrained baseline performance at the point of task switch. If initial performance after the task switch is exactly what it would have been without the initial training, then there is no transfer, and complete specificity. Fig. 1 schematically illustrates full specificity (no transfer), full transfer (no specificity), and a mixed condition of partial specificity and partial transfer. Transfer, or the difference in performance level from an untrained baseline, is marked by vertical blue lines and specificity, or the difference from a fully trained level, is marked by vertical red lines. A specificity *index* (i.e., Ahissar & Hochstein, 1997) often is defined as the proportion of total improvement in the first training stage that is *not* transferred to the new task at the point of the task switch (see Section 2). By analogy, a complementary measure of transfer can also be expressed as the amount (blocks) of training that would have led to the improved perfor-

mance on the new task (see Jeter et al., 2009), illustrated in the green lines dropped to the practice axis.

In this paper, the effects of extended practice on transfer are studied for high precision orientation discrimination for a task switch that changes both the reference angle of the discrimination and the visual location of the test. This relatively high precision task and transfer condition (Jeter et al., 2009) is known to lead to partial transfer and partial specificity after an intermediate level of training. This allows for the measurement of either less or more specificity (or transfer) after different amounts of training.

1.2. Possible frameworks for specificity and transfer

The vast majority of perceptual learning studies train for many sessions and many trials per session, i.e., to asymptotic levels, before testing for transfer to different positions, orientations, or stimuli (Fahle & Poggio, 2002). In this study, we manipulate the extent of training prior to a task switch from very preliminary training to near-asymptotic levels of training over different groups of observers, and observe the extent of transfer to the new task, or extent of specificity of learning to the initial training task.

There are at least four frameworks for understanding specificity, or conversely transfer, in perceptual learning that make predictions for the effect of varying the extent of training.

The dominant framework for understanding specificity is *separate neural representations* (Karni & Sagi, 1991; Gilbert et al., 2001). If the specificity of learning is due to plasticity in early visual cortex for tasks whose different retinal positions or orientations drive relatively independent representations, then specificity should be a consequence of training different cell populations regardless of the amount of training. Specificity occurs because the transfer task uses a new set of previously untrained representations. Under this view, two tasks, say in different locations, will recruit distinct populations in early visual system (i.e., V1) and will show no transfer, or 100% specificity. In this view, specificity is the default, and it is transfer that requires explanation. A modified version could allow a small and constant amount of transfer between the tasks as a consequence of task familiarization.[2] Although this framework is often taken to imply modification of early sensory representations, in fact studies examining transfer between plausibly distinct cortical representations are consistent with either changing sensory representations or reweighting (see below) (see Dosher and Lu (2009) and Petrov, Dosher, and Lu (2005), for a task analysis and review).

An *incremental transfer* framework reasons that something must be learned before it can be transferred, so additional training might lend itself to *more* transfer of a perceptual skill. One can transfer only what is learned, and the incremental transfer is one reasonable corollary. If even some of the improvement during a session transfers, then each added block of training should lead to more improvement – measured as better (absolute) performance on the new task at the point of the task switch. Indeed, at every stage of practice, if any of what is learned in the next block of training is transferred, then performance at the point of transfer should still improve with every incremental block of training. Partial transfer and partial specificity after extensive training is a very common, and perhaps the dominant observation in perceptual learning studies (Crist et al., 1997; Schoups et al., 1995) (see Dosher and Lu (2009) for a review). The ubiquity of partial transfer in the experimental literature is compatible with, but does not directly test, the incremental transfer framework. The incremental transfer prediction is phrased in terms of absolute levels of performance, in which transfer refers to any improvement in second task relative

---

[1] See later in the Introduction for a consideration of baseline issues in the current *matched-task* paradigm.

[2] However, in this experiment the early stages of learning are not easily associated with such factors as learning the key presses or the general experimental environment due to removal of early trials in threshold staircases.

to baseline at the point of the task switch; specific predictions about specificity indexes must be computed for each case.

Another possible framework is the *reverse hierarchy theory* (RHT) of Ahissar and Hochstein (1997, 2004). RHT states that "easy" tasks are learned at higher levels of the visual hierarchy and therefore are transferrable, while "difficult" tasks require learning at lower levels of the visual hierarchy, and are specific to spatial location and for features such as orientation or spatial frequency. More recently, we (Jeter et al., 2009) showed that the operational factor is *not* task difficulty (in the sense of how accurately a task can be performed), but task precision (i.e., the similarity of to-be-discriminated items). The RHT could make one of two predictions about practice of a high precision task (a "difficult" task in their labeling), such as the task used in the current experiment. One prediction of the RHT might be that learning in the "difficult" task in the current experiment is entirely at a low and specific level of the visual hierarchy, and so shows full specificity. Another possible prediction of the RHT is that early improvements reflect changes at high and transferrable levels in the visual hierarchy, while subsequent improvements reflect changes at lower, fully specific, levels. So, the RHT yields the same predictions as the independent neural representation framework, namely, complete specificity with the possible exception of a constant transfer benefit due to the small amount of early and high level learning across all levels of training.

Finally, the *decision optimization* or *reweighting* framework (Dosher & Lu, 1998, 1999; Lu, Chu, Dosher, & Lee, 2005; Petrov, Dosher, & Lu, 2006; Petrov et al., 2005) claims that perceptual learning optimizes connections (decision weights) for a given task by learning incrementally with practice to exclude the least relevant and noisiest information and to up-weight the most relevant and least noisy information. Specificity is a characteristic of the learned connections between early visual representations and task-related decision units, and not a property of the visual representations themselves. The reweighting framework, initially proposed by Dosher and Lu (1998, 1999), was implemented as an Augmented Hebbian Reweighting Model (AHRM) by Petrov et al. (2005, 2006) and experimentally tested with repeated alternation between training phases of discrimination of oriented targets embedded in right-tilted and left-tilted external noise. Continued practice produced both general learning and increased optimization specific to one noise context over the other, seen as persistent switch costs after extended training. To the extent that optimization of weights in the two tasks are not consistent, then training until performance reaches asymptote reinforces learning that is unlikely to transfer, or will transfer negatively to the related task or context. While the transfer between tasks in different retinal locations will require an elaboration of the AHRM computational model for multiple locations, the general principle is that extended practice optimizes specifically for one task and increases switch costs for the transfer task so long as the optimized weight structures for the two differ substantially (see Petrov et al. (2005, 2006) and Dosher and Lu (2009), for reviews). Although the specific predictions depend upon the exact training protocol (see Lu, Liu, and Dosher (2010), for fits of the model to the data of sample experiments), this model predicts early improvements at transfer due to generalized learning, but increased switch costs (lack of transfer, or specificity) after longer periods of training and optimization.

1.3. Experimental approach

Our goal in the current experiment is to measure the amount of transfer, or conversely specificity, following different amounts of training on the initial task. As indicated previously, the paradigm, selected from Jeter et al. (2009), evaluated specificity to a feature (orientation) plus location change between tasks. Modeled on the tasks of Karni and Sagi (1991) and of Ahissar and Hochstein (1997), this paradigm produces intermediate levels of transfer (Jeter et al., 2009), which might then increase or decrease with different levels of practice. Observers trained and tested on a two-alternative high precision (±5° from a fixed, oblique reference angle) identification task with and without external noise masks (see Fig. 2a and b and Section 2). Four groups of observers experience differing amounts of practice prior to transfer, ranging between 2 and 12 blocks. This task showed robust learning in both high and no noise tests and exhibits partial transfer and partial specificity for moderate levels (eight blocks) of training (Jeter et al., 2009), so the extent of transfer after different amounts of training may be measured. There is room to measure both higher and lower specificities as training extent is varied. There is evidence in the literature showing partially independent learning mechanisms with and without external noise masking, so it is possible that the results would differ in the two noise environments (Dosher & Lu, 1998, 2005, 2006; Lu & Dosher, 2009).

There are two approaches to the measurement of transfer in perceptual learning, a matched-tasks method and a pre-test method. The current experiment uses the *matched-task* method in which the initial training task and the transfer task are equivalent and the tasks are randomly assigned to subjects. In this case, the first measurements on the initial training task are the control for performance in the transfer task. The matched-task approach is a good one for the current question because it allows a direct comparison of performance in the two phases. It is also advantageous because there is a clear outcome for 100% specificity – exact equivalence (independence) in the two stages of learning – and a clear outcome for 0% specificity or 100% transfer, in which the performance on the transfer task simply continues that on the training task. The alternative is to pre-test the transfer task, then train on the primary task, and then measure performance on the transfer task after the switch. The pre-testing approach is more complicated to interpret because it requires an estimation of whether the amount of improvement between the first ("pre-test") session to the first session after the transfer switch is larger than some "normal" amount of improvement from a first to a second session (Dosher & Lu, 2007). Additionally, the pre-test approach is called into question by recent work on the enabling of transfer by double training (Xiao et al., 2008), in which transfer may in many cases be specially "promoted" by pre-training some task in the transfer location; in other words a pre-test in the transfer location may set up a special condition in which transfer is more likely to occur. While these double training effects must be understood, and deserve independent study, they complicate the current question. For all these reasons, we selected the matched-task approach to measure specificity and transfer following different amounts of training.

2. Methods

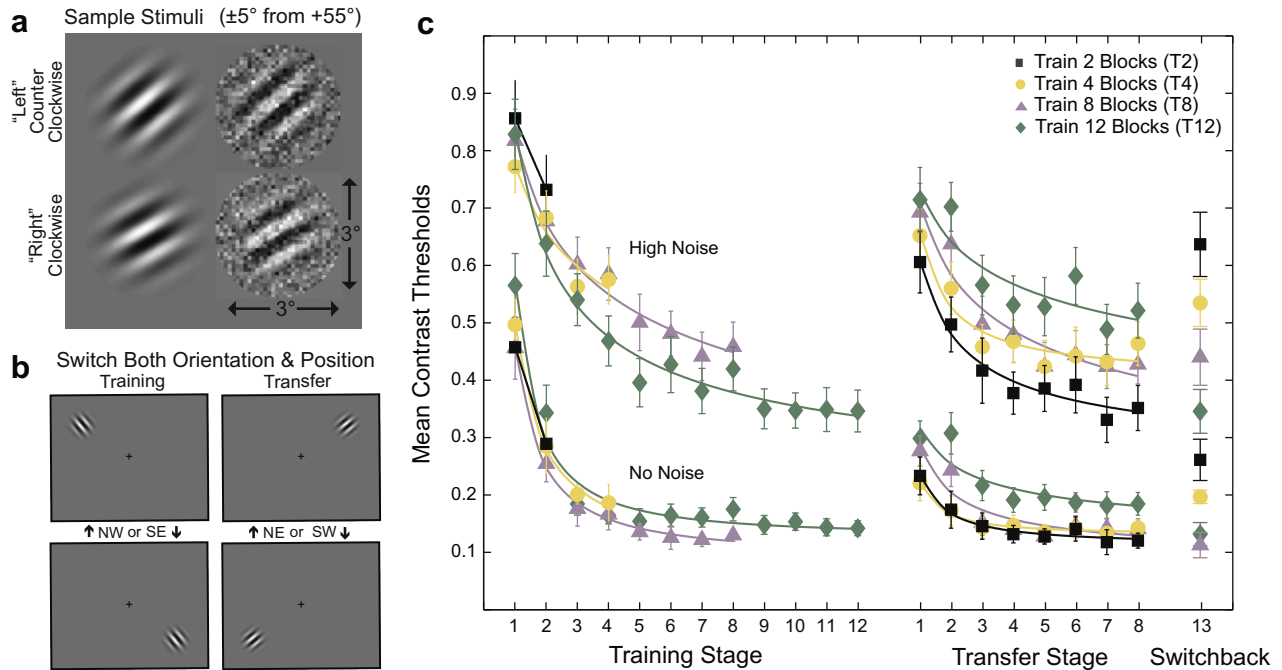### 2.1. Participants

Seven observers participated in each of groups that varied in the amount of initial training of 2, 4, 8, and 12 practice blocks (T2, T4, T8, and T12). All subjects provided written consent under the UC Irvine Institutional Review Board protocol.

2.2. Stimulus and display

The signal Gabor patch was 64 × 64 pixels (3° × 3° visual angle at a viewing distance of 72 cm):

Fig. 2. Sample stimuli, display, and data. (a) Stimuli for a high-precision discrimination task are Gabor targets with and without noise tilted ±5° from an implicit reference angle (+55° shown here, or −35°). (b) Observers trained at one of two pairs on either the NW–SE or the NE–SE diagonal and reference orientation in the training stage and switched both position and orientation in the transfer stage. (c) Average contrast thresholds (75%) during initial training and subsequent practice in the transfer task are shown for conditions trained for either 2, 4, 8, or 12 blocks, in zero noise or in high external noise. High noise trials require higher contrast thresholds than no noise trials. (*Black: Train 2 Blocks* (T2), *Yellow: Train 4 Blocks* (T4), *Purple: Train 8 Blocks* (T8), *Green: Train 12 Blocks* (T12)). All groups practiced for an additional eight blocks in the transfer stage, after switching both positions and angles. The switchback session returned to the original testing conditions. Error bars are two standard deviations estimated using Monte Carlo simulations that resampled from each subject based on the mean and standard deviations of staircase reversals, and averaged over subjects at each data point (resampled 1000 times).

$$l(x, y) = l_0 \left(1.0 \pm c \sin(2\pi f(y \sin(\theta) \pm x \cos(\theta))) \times \exp\left(\frac{x^2 + y^2}{2\sigma^2}\right) \right),$$

with angle θ of $-35° \pm 5°$ or $+55° \pm 5°$, spatial frequency $f = 2$ cpd, standard deviation of the Gaussian envelope $\sigma = 0.4$ degrees. The ±5° angular-difference is a relatively high precision judgment (Jeter et al., 2009). The contrast c is the maximum contrast of the Gabor, and l_0 is the mid-gray luminance. The No Noise and High Noise conditions were intermixed within each testing block. Each 64×64 noise image had individual 2×2 pixel noise elements with Gaussian-distributed values with mean value l_0 and standard deviation 0.33. Signal and noise images were combined via temporal integration (15 ms per frame). Two Gabor frames were 'sandwiched' between pairs of external noise (or blank) frames, so signal and external noise were combined through temporal integration. Fresh noise images were generated for each trial. The stimuli could occur in one of two pairs of retinal positions, either NW/SE or SW/NE corners of the screen, approximately 5.67° of visual angle from fixation (Fig. 2b). On any individual trial, only a single Gabor patch appeared. Each block involved only two diagonally opposite positions. If the first phase of training used the NW/SE diagonal, then the transfer tests used the NE/SW diagonal, and vice versa. All stimuli were generated using MATLAB 5.2 (The Mathworks, 1999) and PsychToolbox 2.34 extensions (Brainard, 1997).

2.3. Apparatus

Stimuli were displayed on a 19″ Viewsonic color monitor by a Macintosh G4 using the internal 10-bit video card (refresh rate 67 Hz, resolution 640 × 480 pixels). Luminance calibration was performed both with psychophysical matching judgments and with a Tectronix Lumacolor J17 photometer. The lookup table di-

vided the luminance range (from 1 cd/m² to 67 cd/m²) into 127 levels for the noise frames and 127 gray levels in the assigned contrast range for the Gabor targets. A chin rest stabilized the observer's head.

2.4. Design

Subjects discriminated between a Gabor tilted clockwise (from top to "Right") or counterclockwise (from top to "Left") from a reference angle of either −35° or +55°. The presentation position was randomized on a diagonal (NW/SE or NE/SW). The reference angle and presentation diagonal were randomly assigned to subjects for initial training (2, 4, 8 or 12 blocks), and switched to the opposite reference angle and diagonal for the transfer tests (all eight blocks). The Gabor orientation discrimination task required high precision judgments of stimuli differing by $\delta° = \pm 5°$ in orientation.

2.5. Procedure

Observers completed 1248 trials per session. Each session was divided into two blocks and separated by brief rest periods. Therefore, T2 trained for 2 blocks (1248 trials in one session in one day), T4 trained for 4 blocks (2496 trials in two sessions over different days), T8 trained for 8 blocks (4992 trials in four sessions over different days) and T12 trained for 12 blocks (7488 trials in six sessions over different days). Contrast thresholds were tracked using adaptive staircases (Levitt, 1971). For the first session in the initial training, transfer and switchback phases, the participant completed 10 practice trials and two practice trials on all other sessions. Additionally, early, 'level-finding', trials (the first 3–5 reversal points or corresponding to 35–60 trials) in each of four interleaved adaptive sequences (see Section 2.6) are not included

in the threshold measurements. On each trial, the participant fixated on a small cross at the center of the screen. A beep occurred 250 ms after the fixation cross. Another 250 ms later, the stimulus sequence (2 white noise frames + 2 Gabor frames + 2 new noise frames) appeared for a total of 90 ms (15 ms/frame). A precue appeared 150 ms after fixation. The short lead-time of 100 ms prior to the oriented Gabor prevented eye movements. A negative feedback tone was presented after each error. The next trial began 750 ms after the key press response.

2.6. Staircase method

Two adaptive staircases (Levitt, 1971) were used to track threshold Gabor contrasts in each stimulus condition. The 3/1 and 2/1 staircases track accuracies of 79.3% and 70.7% correct, respectively. Signal contrast levels were reduced by 10% after either three or two consecutive correct responses and increased 10% after each incorrect response. Separate staircases for all stimulus conditions (including retinal position) were interleaved. There were 168 and 144 trials, respectively, for the 3/1 and 2/1 staircases for a total of 312 in each block. Reversals in staircase direction were determined from the sequence of responses. Threshold contrast levels were computed by averaging an even number of reversals for each staircase sequence, excluding the first four or five. (The number of reversals excluded is either even or odd so as to allow averaging over an even number; this guarantees that every low reversal is balanced with a high reversal, limiting estimation bias.) An overall contrast threshold was estimated by averaging the thresholds of all staircases every two blocks per session (day) in no noise and high noise.

2.7. Methods of analysis

The mean contrast thresholds in different groups were compared with analysis of variance or t-tests or with corresponding non-parametric Kruskal–Wallis or Mann–Whitney U tests. The contrast thresholds as a function of blocks of practice were fitted with power function models, with a lower (minimum threshold) asymptote α and initial incremental threshold λ and a rate ρ: $c(t) = \lambda t^{-\rho} + \alpha$, where t is the number of training blocks. Transfer of perceptual learning at the point of transfer can be measured as the amount of experience that transfers in the context of power function models of perceptual learning: $c(t) = \lambda(t + t_e)^{-\rho} + \alpha$, where the experience parameter t_e summarizes the transfer expressed as the number of blocks of training to yield an equivalent performance (see Dosher & Lu, 2007; Jeter et al., 2009).

In the fullest models, each curve has independent parameters ($\lambda_i \, \rho_i \, \alpha_i$, for each condition i), while constrained (nested) models held certain parameters equal (equating $\lambda_i = \lambda_j$, $\rho_i = \rho_j$, and/or $\alpha_i = \alpha_j$ over groups). If groups differ by one or more parameters, then a nested model comparison equating those parameter(s) will significantly reduce the quality of fit when tested by a nested-model F:

$$F(df1, df2) = \frac{\left(r_{fuller}^2 - r_{less-full}^2\right)/(k_{fuller} - k_{less-full})}{\left(1 - r_{fuller}^2\right)/(n - k_{fuller})},$$

$$df1 = (k_{fuller} - k_{less-full}), \quad df2 = (n - k_{fuller}).$$

In some cases, the p-values of several parametric or non-parametric tests are combined through the Fisher's $\chi^2 = -2\sum_{i=1}^{k} \ln p_i$, where k is the number of significance tests, and $df = 2k$ (Fisher, 1932).

Specificity can be measured in several ways. The specificity index introduced by Ahissar and Hochstein (1997) expresses specificity as the proportion of the improvement during initial

training that does *not* transfer, a score that expresses specificity as a percentage in the dimension of contrast threshold: $S_c = \left(C_{X1}^i - C_{Tend}^i\right)/\left(C_{T1}^i - C_{Tend}^i\right)$, where C_{T1}^i and C_{Tend}^i are the contrast thresholds for the first and last blocks of the initial training phase and C_{X1}^i is the contrast threshold for the first block of the transfer test, where i indicates the group or condition. Ahissar and Hochstein's specificity index is best suited where initial training is asymptotic. The current experiment measures transfer after small amounts of practice in certain conditions (T2, T4), and in these situations the index can be improved by replacing C_{Tend}^i with C_{Tend+1}^i – the value where the contrast threshold is expected to be at the next testing block under full transfer. This index is: $S_c' = \left(C_{X1}^i - C_{Tend+1}^i\right)/\left(C_{T1}^i - C_{Tend+1}^i\right)$. Alternatively, the transfer value t_e from the power function models of perceptual learning provides a measure of transfer in the dimension of practice blocks; a t_e of 0 corresponds to no transfer, of full specificity, while $t_e = t_{Ttotal}$ corresponds to full transfer, or no specificity.

3. Results

3.1. Learning curves for initial training

Four groups received different amounts of 2, 4, 8, or 12 blocks of practice before transfer, labeled T2, T4, T8, and T12, respectively. Observers are assigned at random to these groups that vary only in the amount of initial training. After training, all groups switched to the opposite retinal positions and orientation and completed eight additional blocks of practice on the transfer task. Two blocks of training occurred per day. Task performance is indexed by the average (threshold) contrast required to produce a criterion accuracy level (i.e., 75% correct) by averaging two adaptive staircases, 70.7% for a 2/1 staircase, and 79.3 for a 3/1 staircase (see Methods).

Group average contrast thresholds (at 75% correct) were plotted as a function of practice block separately for 'no noise' and 'high noise' test conditions (Fig. 1c, left half). The smooth curves in Fig. 2 are fits of a power function model of improvement in contrast threshold as a function of training block: with a lower (minimum threshold) asymptote α and initial incremental threshold: $c(t) = \lambda t^{-\rho} + \alpha$, applied independently to the initial training and transfer phases for each group separately. Explicit estimation of transfer within a joint power function model appears in Section 3.3. Thresholds were higher with external noise masks than without (all $p < 0.001$). All curves showed significant reductions in thresholds between the first and last block in the training phase, with more training producing larger contrast threshold reductions (paired t-tests, all $p < 0.03$).

Although the *total* amount of learning should increase with increased training, we expected no differences in the common (overlapping) portions of the training performance as the four groups differ only in random assignment of observers. The first two blocks of all groups (T2, T4, T8, T12) showed no differences in either the high external noise or no noise conditions (all individual $p > 0.46$ in Kruskal–Wallis tests), with Fisher's $\chi^2 = 2.26$, $df = 4$, $p = 0.688$ for high noise and Fisher's $\chi^2 = 1.09$, $df = 4$, $p = 0.896$ for low noise. The first eight blocks of T8 and T12 did not differ significantly (all individual $p > 0.23$ in high noise and >0.18 in no noise in Mann–Whitney tests), with Fisher's $\chi^2 = 14.14$, $df = 16$, $p = 0.588$ for high noise and Fisher's $\chi^2 = 17.41$, $df = 16$, $p = 0.360$ for low noise.[3] In short, more training produces more learning, but the groups – as

---

[3] Non-parametric tests were requested by reviewers in consideration of the somewhat smaller number of observers in each group; these results were closely equivalent to the corresponding parametric tests.
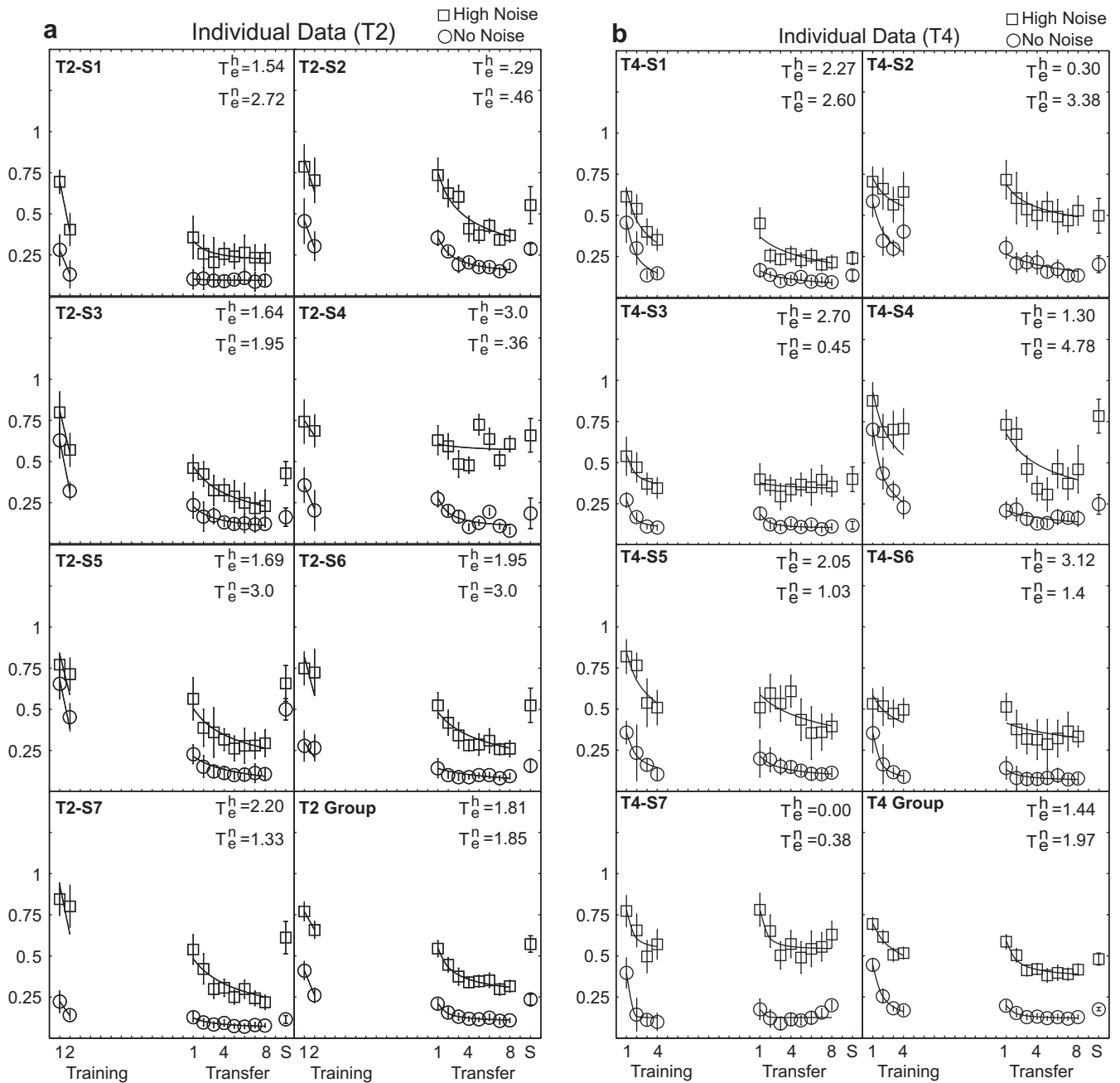
Fig. 3. Perceptual learning data in high and no noise for individual data and the group average. Contrast thresholds are plotted for the seven individuals in each group and the group average data: (a) T2, (b) T4, (c) T8, and (d) T12. The smooth curves are the best-fitting power function model with experience, or transfer parameter, t_e, free to vary shown in Table 1. See the text and Table 2 for comparisons with other nested models.

expected – were statistically equivalent for the shared portions of the training curves.

3.2. Transfer phase performance

The key question of this research is whether the performance on the transfer test depends upon the amount of initial training. Group T2 yielded the best average performance on the transfer task, while T12 showed the worst average performance on the transfer test, and so the least gain given the amount of training, with Groups T4 and T8 showing intermediate values (Fig. 2c). Comparing the shortest and longest training groups (T12 versus T2) with Kruskal–Wallis tests, T12 showed poorer transfer performance in high external noise (Fisher's $\chi^2 = 40.37$, $df = 16$,

$p < 0.001$; ps for the eight transfer blocks [0.025–0.338]) and in no noise (Fisher's $\chi^2 = 37.33$, $df = 16$, $p < 0.001$; ps for the eight transfer blocks [0.035–0.227]). (The corresponding parametric analysis of variance tests were significant for high external noise and just missed significance in no noise.)

Power function models of the average group data during the transfer phase also documented worse transfer performance in T12 (contrast thresholds were higher) than in the T2 training group. A model that restricted the power function learning curves in the training phase to be the same (equating $\lambda_2 = \lambda_{12}$, $\rho_2 = \rho_{12}$, and $\alpha_2 = \alpha_{12}$) was easily rejected in both high noise ($F(3, 10) = 26.68$, $p = 0.001$, for the nested model test) and no noise ($F(3, 10) = 16.75$, $p = 0.001$), again indicating poorer performance at transfer following more training.
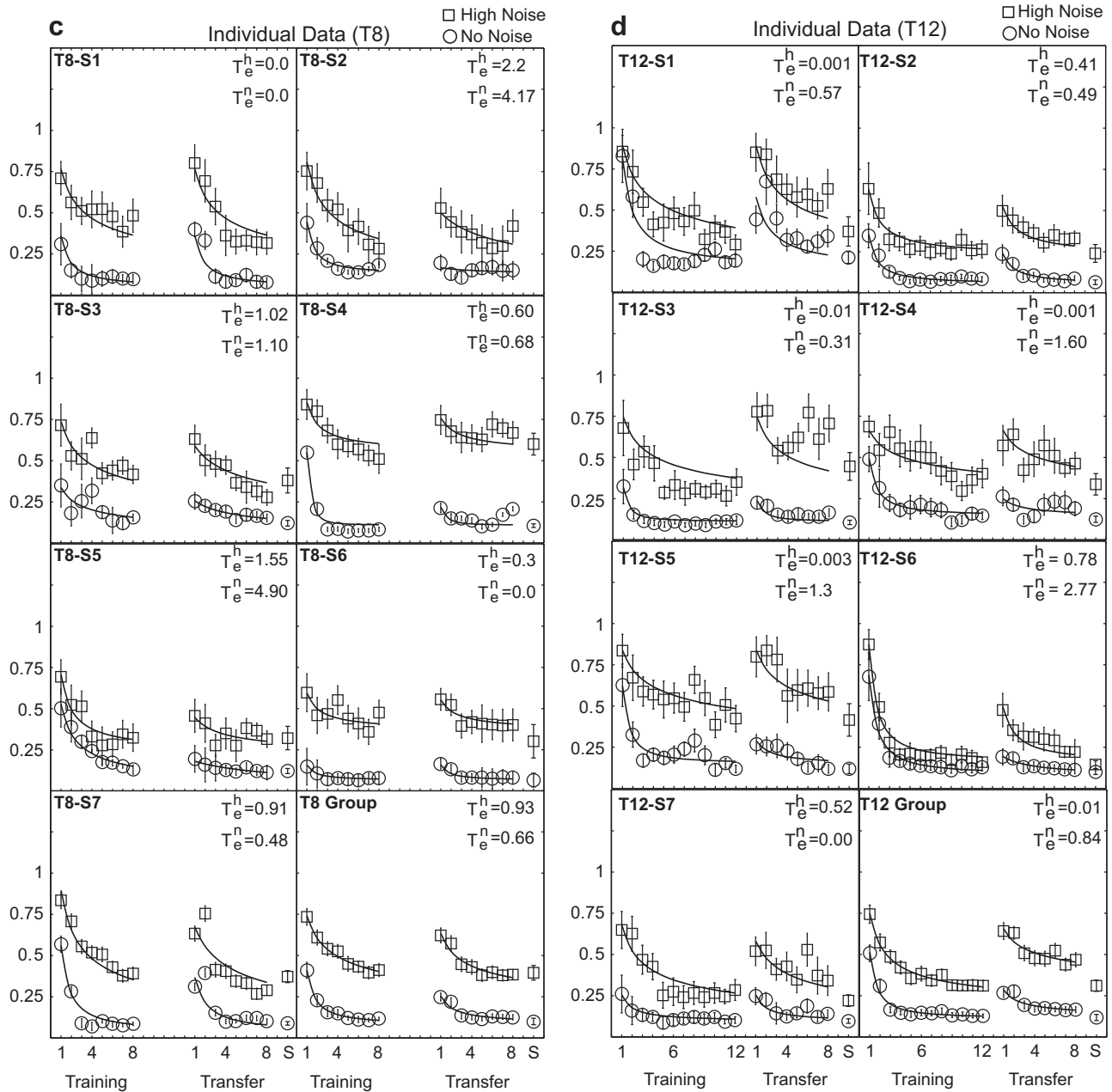
Fig. 3 (*continued*)

By all these measures on the aggregate data, more training yielded significantly less transfer (worse contrast threshold performance) after the task switch than did brief training.

3.3. Power functions and T_e

This section uses expanded power function models that explicitly estimate a transfer or experience parameter after different amounts of practice. This approach has advantages: (1) It derives a model-based estimate of transfer, and (2) It can be applied and tested for significance in individual as well as group data. The initial training phase and the transfer phase are taken together, with parameter t_e estimating the benefit from prior training in the transfer phase – measured in blocks of training required to match performance at the point of initial transfer (see Dosher & Lu, 2007; Jeter et al., 2009). The initial training and transfer phases are fit with common asymptote α, initial incremental threshold λ, and a rate ρ – reflecting the structure of the matched-task design. The value of t_e is set to 0 in training and is estimated from the data for the transfer phase.[4]

Individual data are shown in Fig. 3a–d, with separate panels for each training group. Table 1 lists the estimated values of the experience parameter t_e, and R^2 summarizing the quality of fit for high external noise and no noise for each individual and the group average. While the threshold levels and rate of learning differ between subjects (individual learning is notoriously variable in perceptual learning studies, Fine & Jacobs, 2002), there is a relatively

---

[4] Subsidiary analyses showed separate rates of learning ρ in the initial training and transfer phases when compared directly. Whenever t_e is greater than 0, the elaborated power function itself embodies a slower (instantaneous) rate of learning due to the transfer phase starting farther along the power function.

Table 1
Estimated t_e transfer scores for primary model.

Primary model with t_e free to vary											
Group T2			Group T4			Group T8			Group T12		
Subject	t_e	R^2	Subject	t_e	R^2	Subject	t_e	R^2	Subject	t_e	R^2
No noise											
T2-S1	2.718	0.986	T4-S1	2.601	0.954	T8-S1	0.000	0.796	T12-S1	0.569	0.672
T2-S2	0.464	0.978	T4-S2	3.379	0.875	T8-S2	4.173	0.910	T12-S2	0.492	0.940
T2-S3	1.947	0.996	T4-S3	0.454	0.945	T8-S3	1.101	0.605	T12-S3	0.313	0.803
T2-S4	0.359	0.873	T4-S4	4.777	0.981	T8-S4	0.679	0.881	T12-S4	1.602	0.810
T2-S5	3.000	0.974	T4-S5	1.029	0.944	T8-S5	4.894	0.971	T12-S5	1.300	0.835
T2-S6	3.000	0.883	T4-S6	1.412	0.978	T8-S6	0.000	0.870	T12-S6	2.769	0.976
T2-S7	1.330	0.981	T4-S7	0.380	0.874	T8-S7	0.475	0.829	T12-S7	0.000	0.764
Group	1.850	0.990	Group	1.966	0.992	Group	0.659	0.969	Group	0.841	0.928
High noise											
T2-S1	1.540	0.970	T4-S1	2.265	0.886	T8-S1	0.000	0.764	T12-S1	0.001	0.748
T2-S2	0.291	0.908	T4-S2	0.304	0.773	T8-S2	2.215	0.862	T12-S2	0.405	0.867
T2-S3	1.645	0.990	T4-S3	2.696	0.762	T8-S3	1.023	0.722	T12-S3	0.000	0.360
T2-S4	3.000	0.340	T4-S4	1.298	0.722	T8-S4	0.596	0.575	T12-S4	0.001	0.566
T2-S5	1.693	0.905	T4-S5	2.054	0.783	T8-S5	1.549	0.798	T12-S5	0.003	0.700
T2-S6	1.945	0.898	T4-S6	3.124	0.678	T8-S6	0.302	0.617	T12-S6	0.775	0.943
T2-S7	2.198	0.898	T4-S7	0.000	0.758	T8-S7	0.908	0.823	T12-S7	0.521	0.711
Group	1.810	0.960	Group	1.443	0.909	Group	0.930	0.948	Group	0.006	0.824

Note: This table reports the estimated experience or transfer factors t_e and the quality of fit, R^2, of the extended power function model: $c(t) = \lambda(t + t_e)^{-\rho} + \alpha$, where $t_e \equiv 0$ for initial training, and t_e estimated for the transfer phase. For the current matched-task design, $\alpha_T = \alpha_X = \alpha$, $\lambda_T = \lambda_X = \lambda_T$, and $\rho_T = \rho_X = \rho$, with subscripts T for initial training and X for transfer. This model assumes that the two phases are identical except for transfer factor t_e.

consistent pattern of increasing specificity with added training. Values of t_e show a downward trend or less transfer for groups receiving more training (see Table 1). The average t_e's in high noise were 1.81, 1.44, 0.93, and 0.01, for T2, T4, T8, and T12, respectively ($\chi^2 = 9.26$, $df = 3$, $p = 0.03$), and in low noise were 1.85, 1.96, 0.66, and 0.84, respectively ($\chi^2 = 1.79$, $df = 3$, $p = 0.618$ by non-parametric Kruskal–Wallis tests). Considering comparisons of T2 and T12: In *high noise* conditions, the average $t_e = 1.81$ for T2 and $t_e = 0.01$ for T12 differed significantly ($z = -2.75$, $p = 0.006$ by Mann–Whitney test). In *no noise* conditions, $t_e = 1.85$ for T2 and $t_e = 0.84$ in T12 ($z = -1.34$, $p = 0.18$). In short, T12 shows significantly worse performance at the point of transfer than T2. Indeed, the average estimated transfer decreases systematically from near full transfer at T2 towards little transfer at T12.

3.4. Nested model tests for T2 and T12 groups

The results of individual observers, as well as the average, can be tested for significant differences from 100% specificity ($t_e = 0$) or 100% transfer (t_e = number of initial practice blocks) through nested model test comparison with models in which t_e is an estimated (free) parameter. Full specificity ($t_e = 0$) is consistent with the independent representations framework as a consequence of training and testing different neural representations regardless of the amount of training. For the T2 group, the constrained model (Table 2, *bottom*), where $t_e = 0$ (100% specificity) was rejected in the average data in both high external noise ($F(1, 6) = 31.65$, $p = 0.001$) and no noise ($F(1, 6) = 185.27$, $p = 0.0001$). It was rejected for 5 of 7 individuals in high noise and for 6 or 7 individuals in low noise conditions. In the cases where the 100% specificity model was rejected, with freely estimated $t_e > 0$, indicating significant transfer. For the T2 group, setting $t_e = 2$ (the number of blocks of initial training in T2) was statistically equivalent to the model with t_e free to vary in both high external noise ($F(1, 6) = .087$, $p = 0.778$) and in no noise ($F(1, 6) = 0.277$, $p = 0.618$) for the average data (see Table 2, *top*) and for the majority of individual observers (6 of 7 observers in high noise and 5 of 7 observers in low

no noise). In short, for the majority of observers and for the group data, two blocks of training yielded results that are statistically consistent with 100% transfer.

In contrast, full transfer (setting $t_e = 12$) was easily rejected ($p < 0.001$) for the T12 condition for the average data in both high noise ($F(1, 12) = 43.54$, $p < 0.001$) and no noise ($F(1, 12) = 38.41$, $p < 0.001$), and for all individual observers. Instead, the results for T12 are closer to 100% specificity. In high noise, the average data showed no significant difference from 100% specificity, or no transfer ($t_e = 0$) ($F(1, 12) = 0.005$, $p = .942$), with 1 (and one marginal) of 7 observers showing significantly more transfer. In low noise, although the estimated transfer scores were all less than 1, these small levels of transfer were significantly different from zero for the average data ($F(1, 12) = 37.62$, $p < 0.001$), and 6 of 7 observers. Still, the extent of transfer is significantly less than $t_e = 2$, which was rejected in high noise in the average data ($F(1, 12) = 17.40$, $p = 0.001$) and for 6 of 7 individual observers, and was rejected in no noise for the average data ($F(1, 12) = 10.91$, $p = 0.005$), and for 5 of 7 individual observers.

The most extended training group (T12) approximated 100% specificity, especially in high noise – which might suggest complete independence of learning in the transfer stage from the perceptual learning in the initial training stage. If so, the eight blocks of training after the task switch should equal the first eight blocks of initial training in the same condition of this matched-task design. However, the data suggest that the rate of perceptual learning in the transfer task for group T12 may be slowed in both no noise and high noise. The power function model where all parameters are free to vary for the first eight blocks in training in T12, ($\rho = -1.35$ in No Noise and $\rho = -0.61$ High Noise) and during the transfer stage for all eight blocks in T12 ($\rho = -0.28$ in No Noise and $\rho = -0.24$ High Noise) are compared in a separate nested model test that equates ρ (rate) and α. The first eight blocks in the initial training differ from those in the transfer training for this condition ($F(2, 10) = 5.94$, $p = 0.02$) in no noise, and ($F(2, 10) = 10.33$, $p = 0.003$) in high noise indicating that the difference in rates between initial training and transfer differed significantly. This suggests a more complicated interpretation for

Table 2
Nested-significance tests for several models of t_e.

Group T2				Group T4				Group T8				Group T12			
Subject	R^2	$F(1, 6)$	ρ	Subject	R^2	$F(1, 8)$	ρ	Subject	R^2	$F(1, 12)$	ρ	Subject	R^2	$F(1, 12)$	ρ
T_e = Fixed # of training blocks															
No noise															
T2-S1	0.985	0.564	0.481	T4-S1	0.947	1.167	0.311	T8-S1	0.178	36.338	0.001*	T12-S1	0.455	10.438	0.005*
T2-S2	0.893	23.295	0.003*	T4-S2	0.874	0.093	0.769	T8-S2	0.907	0.436	0.522	T12-S2	0.647	78.101	0.000*
T2S3	0.996	1.186	0.318	T4-S3	0.7S6	29.401	0.001*	T3-S3	0.424	5.732	0.034*	T12-S3	0.576	18.440	0.001*
T2-S4	0.763	4.947	0.068	T4-S4	0.980	0.020	0.890	T8-S4	0.829	5.243	0.041*	T12-S4	0.747	5.323	0.035*
T2-SS	0.958	3.619	0.106	T4-SS	0.829	16.412	0.004*	T8-S5	0.957	5.363	0.039*	T12-S5	0.767	7.259	0.016*
T2-S6	0.847	2.323	0.178	T4-S6	0.954	10.404	0.012*	T3-S6	0.250	57.276	0.000*	T12 S6	0.950	17.460	0.001*
T2-S7	0.975	1.556	0.259	T4-S7	0.845	1.832	0.213	T8-S7	0.577	17.834	0.001*	T12-S7	0.312	30.460	0.000*
Group	0.990	0.277	0.618	Group	0.974	12.574	0.008*	Group	0.754	86.482	0.000	Group	0.755	38 413	0.000
High noise															
T2-S1	0.968	0.324	0.590	T4-S1	0.871	1.040	0.338	T3-S1	0.274	24.898	0.001*	T12-S1	0.193	35.621	0.000*
T2-S2	0.778	6.517	0.043*	T4-S2	0.581	6.753	0.032*	TS-S2	0.744	10.239	0.008*	T12-S2	0.514	42.454	0.000*
T2-S3	0.987	1.549	0.260	T4-S3	0.759	0.032	0.862	T8-S3	0.524	8.392	0.013*	T12-S3	0.053	7.680	0.014*
T2-S4	0.333	0.067	0.804	T4-S4	0.676	1.256	0.295	TS-S4	0.379	5.543	0.036*	T12-S4	0.294	10.262	0.006*
T2-S5	0.904	0.061	0.813	T4-SS	0.751	1.186	0.308	T8-SS	0.701	5.934	0.031*	T12-S5	0.196	26.905	0.000*
T2-S6	0.898	0.121	0.740	T4-S6	0.677	0.084	0.780	T8-S6	0.357	8.307	0.014*	T12-S6	0.756	52.571	0.000*
T2-S7	0.897	0.180	0.687	T4-S7	0.323	14.383	0.005*	T8-S7	0.560	17.769	0.001*	T12-S7	0.340	20.550	0.000*
Group	0.959	0.087	0.778	Group	0.844	5.901	0.041*	Group	0.625	77.994	0.000	Group	0.345	43.544	0.000
T_e = 2, Fixed															
No noise															
T2-S1	0.985	0.564	0.481	T4-S1	0.951	0.608	0.458	T3-S1	0.348	26.367	0.001*	T1 2-S1	0.579	4.539	0.049*
T2-S2	0.893	23.295	0.003*	T4-S2	0.851	0.905	0.369	T8-S2	0.898	1.688	0.213	T12-S2	0.843	25.846	0.001*
T2-S3	0.996	1.183	0.318	T4-S3	0.837	15.701	0.004*	T8-S3	0.586	0.574	0.463	T12-S3	0.628	13.774	0.002*
T2-S4	0.763	4.947	0.068*	T4-S4	0.949	13.616	0.006*	T3-S4	0.847	3.446	0.088	T12-S4	0.808	0.151	0.703
T2-S5	0.958	3.619	0.106	T4-SS	0.910	4.394	0.058*	T8-SS	0.913	22.738	0.001*	T12-S5	0.827	1.306	0.270
T2-S6	0.847	2.323	0.178	T4-S6	0.974	1.623	0.239	T8-S6	0.427	40.894	0.000*	T12-S6	0.972	6.179	0.024*
T2-S7	0.976	1.556	0.259	T4-S7	0.847	1.707	0.228	T3-S7	0.697	9.276	0.010*	T12-S7	0.471	19.244	0.001*
Group	0.990	0.277	0.618	Group	0.992	0.401	0.544	Group	0.892	29.948	0.000	Group	0.879	10.905	0.005
High noise															
T2-S1	0.968	0.324	0.590	T4-S1	0.885	0.387	0.551	T8-S1	0.541	11.317	0.006*	T12-S1	0.476	17.280	0.001*
T2-S2	0.778	6.517	0.043*	T4-S2	0.685	3.115	0.116	T8-S2	0.861	0.077	0.786	T12-S2	0.732	16.231	0.001*
T2-S3	0.987	1.549	0.260	T4-S3	0.760	0.085	0.778	T3-S3	0.701	0.922	0.356	T12-S3	0.134	5.650	0.030*
T2-S4	0.333	0.067	0.804	T4-S4	0.715	0.210	0.659	TS-S4	0.504	1 891	0.134	T12-S4	0.585	0.016	0.900
T2-S6	0.904	0.061	0.313	T4-S6	0.783	0.003	0.961	T8-S6	0.794	0.247	0.623	T12-S5	0.468	12.386	0.003*
T2-S6	0.898	0 121	0.740	T4-S6	0.672	0.144	0.714	T3-S6	0.556	2.023	0.180	T12-S6	0.887	14.085	0.002*
T2-S7	0.897	0 180	0.687	T4-S7	0.417	11.251	0.010*	TS-S7	0.786	2.487	0.141	T12-S7	0.630	4.503	0.050*
Group	0.959	0.087	0.778	Group	0.902	0.637	0.448	Group	0.906	9.799	0.009*	Group	0.633	17.395	0.001
T_e = 0, Fixed															
No noise															
T2-S1	0.447	222.858	0.000*	T4-S1	0.525*	4.517	0.000*	T8-S1	0.796	0.239	0.634	T12-S1	0.623	2.373	0.143
T2-S2	0.909	18736	0.005*	T4-S2	0.527	22.225	0.002*	T8-S2	0.478	57 802	0.000*	T12-S2	0.866	19.774	0.000*
T2-S3	0.606	651.323	0.000*	T4-S3	0.806	19.600	0.002*	T8-S3	0.530	2.266	0.158	T12-S3	0.749	4.415	0.052
T2-S4	0.823	2.350	0.176	T4-S4	0.448	223.420	0.000*	T8-S4	0.603	23.018	0.000*	T12-S4	0.567	20.497	0.000*
T2-S5	0.542	99.731	0.000*	T4-SS	0.738	29.731	0.001*	T8-SS	0.467	201.194	0.000	T12-S5	0.577	24.990	0.000*
T2-S6	0.463	21.493	0.004*	T4-S6	0.610	130.662	0.000*	T8-S6	0.870	0.055	0.819	T12-S6	0.552	282.981	0.000
T2-S7	0.712	34.154	0.000*	T4-ST	0.554	20.251	0.002*	T8-S7	0.760	4.805	0.049*	T12-S7	0.764	0.031	0.863
Group	0.681	185.274	0.000*	Group	0.598	393.931	0.000*	Group	0.830	53.650	0.000	Group	0.759	37.622	0.000
High noise															
T2 S1	0.623	69.699	0.000*	T4-S1	0.646	16.732	0.004*	T3-S1	0.764	0.002	0.965	T12-S1	0.748	0.023	0.882
T2-S2	0.894	0.955	0.366	T4-S2	0.763	0.338	0.577	T8-S2	0.623	20.763	0.001*	T12-S2	0.833	4.108	0.060
T2-S3	0.751	140.660	0.000*	T4-S3	0.449	10.477	0.012*	T8-S3	0.658	2.778	0.121	T12-S3	0.360	0.001	0.980
T2-S4	0.142	1.803	0.228	T4-S4	0.619	2.963	0.124	T8-S4	0.532	1.206	0.294	T12-S4	0.566	0.136	0.717
T2-S5	0.722	11.569	0.015*	T4-S5	0.540	9.007	0.017*	T8-S5	0.588	12.422	0.004*	T12-S5	0.700	0.019	0.892
T2-S6	0.690	12.204	0.013*	T4-S6	0.521	3.911	0.083	T3-S6	0.604	0.427	0.526	T12-S6	0.790	43.033	0.000*
T2-S7	0.665	13.601	0.010*	T4-S7	0.758	0.072	0.795	T8-S7	0.743	5.431	0.038*	T12-S7	0.687	1.361	0.260
Group	0.749	31.654	0.001*	Group	0.759	13.159	0.007*	Group	0.865	19.135	0.001	Group	0.824	0.005	0.942

Note: Results of nested significance tests of nested models that fix t_e, reporting the R^2 summary of goodness of fit for the fixed model, the F-test for significance relative to the fuller model with t_e free to vary, reported in Table 1, and the corresponding p-value. One model "t_e = 0" assumes 0% transfer, corresponding to 100% specificity. Another model, "t_e = Number of training blocks" assumes 100% transfer, or 0% specificity of training. A third model, "t_e = 2" assumes transfer of the first two training blocks (1st day) only.
* Indicates significant differences between the free and constrained model, at least at the 0.05 level.

specificity than merely training and testing independent neural representations.

Taken together, the nested model tests on t_e suggest that after additional training until asymptote, performance is worse in the transfer stage for group T12 than for the T2 group that trained

the least, thus reinforcing the notion that optimized learning is less likely to transfer. The majority of F-tests for the individual observers were generally consistent with these conclusions (Fig. 3a–d). The T4 and T8 groups having an intermediate pattern of specificity, suggesting a continuous reversal of transfer towards specificity

with increased training.[5] These results are remarkable – improvements seen after transfer to a second task following only two blocks of training are largely eliminated after more extended training to yield increasing specificity.

A framework that suggests specificity is due to training and testing separate neural populations in early visual cortex cannot account for the data. A framework suggesting that more training can lead to additional transfer is also rejected. We are lead to infer that learning is a dynamic process that has different consequences for transfer and specificity for different amounts of training. Possible interpretations are considered in the discussion.

3.5. Specificity index

The pattern described in the previous sections has a parallel expression when summarized with specificity indices. Specificity indices quantify the performance at the initial point of transfer as a proportion of the total improvement in the training phase that does *not* transfer. An initial performance in the transfer phase that matches the initial performance in the first task (assuming equivalence of the two tasks) corresponds to a specificity of 1.0. Fig. 4 shows the results for a form of the index $S'_c = \left(C^i_{X1} - C^i_{Tend+1} \right) / \left(C^i_{T1} - C^i_{Tend+1} \right)$ (see Section 2 for discussion) that uses the final contrast threshold expected at the next block C^i_{Tend+1} as the estimated final learning in the training stage in order to take into account rapid learning in the early stages of practice. This is modified from the standard index (Ahissar & Hochstein, 1997; Jeter et al., 2009), developed for cases were initial training has already reached asymptotic levels of performance, which omits the +1 in the subscript. (The standard index leads slightly smaller estimates especially for briefer training conditions, and yields a negative estimate of specificity for T2, corresponding to performance even better than the final block of the initial task.) Specificity indices are plotted for each training group for tests with and without external noise masks (Fig. 4). The specificity scores increase with the amount of initial training, with the highest specificity for T8 and T12. These specificity indices are generally also higher for high noise test conditions than for no noise tests. In high noise trials, all groups except T2 showed significant degrees of specificity as measured by the index (all $p < 0.05$ by one sample t-test), while those for T2 were not significantly different from zero (n.s.). In no noise trials, T2 and T4 specificity scores showed no significant difference from zero, indicating transfer (all $p > 0.10$) while T8 and T12 showed partial specificity. The high noise tests may have larger specificity indices because these tests are more sensitive to mistuning of weight templates for external noise exclusion (Dosher & Lu, 1998) upon changing to the transfer task.

3.6. Switchback session

We also evaluated the impact of continued training during the second, transfer phase (the eight blocks after the task switch) on the performance of the initial task through a "switchback" test. A subset of subjects was switched back to the initial training task in an additional session after training and transfer stages were completed (see Fig. 2c right and Section 3.6). The performance in the switchback block is statistically equivalent, with one exception, to the last block in the initial training phase (see Fig. 5), which
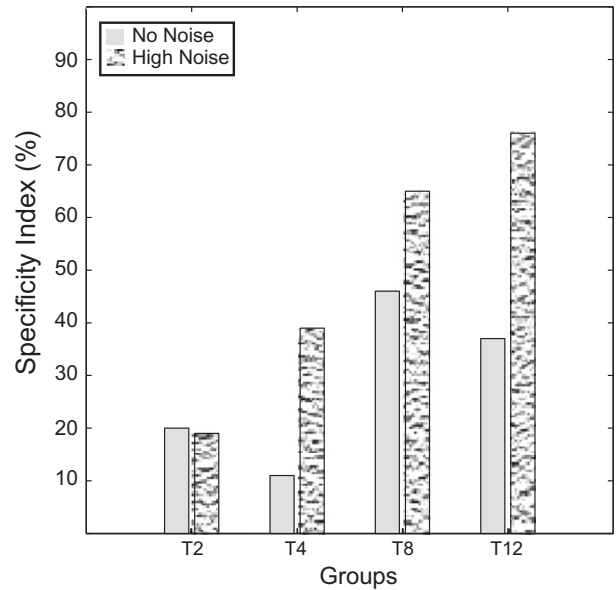

Fig. 4. Specificity Index for aggregate data in high and no external noise. The specificity index $S'_c = \left(C^i_{T1} - C^i_{Tend+1} \right) / \left(C^i_{X1} - C^i_{Tend+1} \right)$ takes into account the rapid improvements in early learning for short training groups. The index is shown for high external noise (pixilated) and no external noise (gray) test conditions. Specificity systematically increases with the amount of training, and is larger in external noise tests.
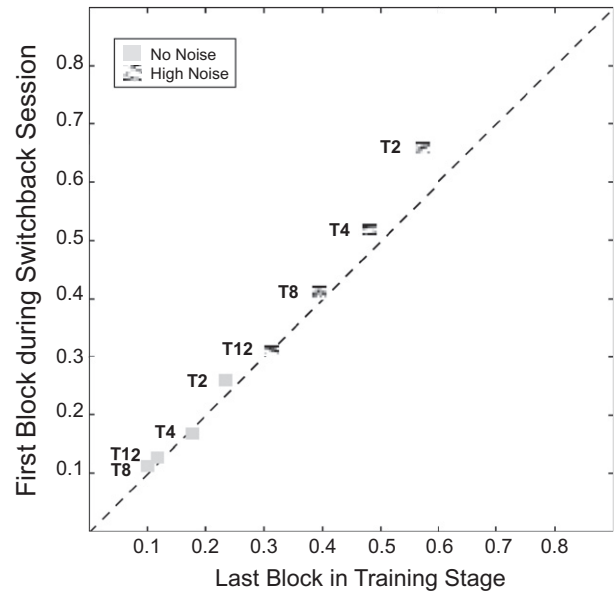

Fig. 5. Analysis of switch-back performance. The contrast threshold for a switchback test following training in the transfer task is compared to the last contrast threshold of the same task in the initial training phase. The two are nearly equivalent, except for T2, which shows additional learning.

tested the identical task, stimuli, and retinal location (all $p > 0.10$ by paired t-test with the exception of T2 in Hi Noise where the switchback threshold is significantly better i.e., lower than the threshold from the last day of initial training). Despite having practiced for eight additional blocks on a different position and orientation during the transfer phase, performance on the initial task was essentially unchanged. There are two possible interpretations of these data. One is that there is an asymmetry of influence preserving the earlier-learned information: training in the initial stage task alters performance in the transfer stage task, but training in

---

[5] An anonymous reviewer suggested that individual observers show either full transfer or full specificity, with different mixtures in different training conditions. Due to the fact that each observer appears in only one group, we cannot rule out this interpretation. However, continuous changes in optimized weight structures provide an alternative and consistent account of the results, one we feel is more consistent with the overall pattern of data.

the transfer phase task does not go back to corrupt learning on the initial training task. Another possibility – one that we favor – is that there are two influences on learning, namely general learning improvements, and specific learning switch costs, that oppose and approximately cancel one another in this situation. These possibilities are considered further below.

4. Discussion

In summary, the group that trained the least (T2), corresponding to the early stages of the training process where learning is most rapid (Hawkey, Amitay, & Moore, 2004; Poggio, Fahle, & Edelman, 1992), also has the most transfer. While providing the best final performance for the first task, training that approaches asymptotic performance (e.g., T12) also engages specific learning that increasingly limits the transfer to the second task with similar but different stimuli and judgments. Groups that have not reached asymptotic performance in training (e.g. T4, T8) show an intermediate pattern of transfer at the initial point in the transfer stage. Continued training on the transfer task results in perceptual learning of that task. So, the most specificity, or least transfer, was observed for group T12 with the most training at the initial point of transfer. Although the T12 performance in the transfer task with different orientation and locations showed almost full specificity at the switch point, yet there was some indication in the data that the subsequent perceptual learning of the new task is not quite as efficient as the initial learning: The learning in the transfer task was slightly but significantly slower than the learning in the original task for the same number of training blocks. When switched back to the initial training task in the final session, performance was essentially where it left off. So either the intervening training on the transfer task does not interfere significantly with the original learning, or else, as we believe, increasing specificity offsets ongoing general learning[6]. These results demonstrate that specificity of learning to stimulus dimensions such as orientation and retinal position changes dynamically over the course of training. To our knowledge, this is the first systematic empirical examination of the effects of increased training on transfer in perceptual learning. The current study examined the case of transfer to a different feature value (i.e., reference orientation angle) in different visual location, which plausibly incurs switch costs due to the inconsistency in orientation angles. Further research is needed to fully understand the boundary conditions for the phenomena.

In any case, the results of the current experiment would seem to rule out several of the hypotheses about transfer and specificity outlined in the introduction. The *separate neural representations framework* in its simple form predicts specificity regardless of the extent of initial training since the two tasks are assumed to train separate and independent representations for the different orientations and different visual positions. That initially transferrable improvements are eliminated and reversed with extended training suggests that the classic observations of specificity following asymptotic training does not reflect retuning or other modification of different, independent, pieces of visual cortex (Ahissar & Hochstein, 1997; Gilbert et al., 2001; Karni & Sagi, 1991), at least in any simple way. Even if this framework were modified to allow for some small amount of general learning at the beginning, these initial benefits due to general task familiarization should be maintained even if subsequent learning is 100% specific.

The *incremental transfer* framework argues that whatever is acquired at each incremental stage of learning has some chance of transfer, so that net transfer can only be improved with further training. This is also inconsistent with our findings – it predicts exactly the opposite ordering of empirical transfer reported here, such that the longest initial training should have had the highest, not the lowest amount of benefit at the point of task switch.

The *reverse hierarchy theory* (Ahissar & Hochstein, 1997, 2004) states that "easy" tasks are learned at higher levels of the visual hierarchy and therefore are transferrable, while "difficult" tasks require learning at lower levels of the visual hierarchy, and are specific to spatial location and for features such as orientation or spatial frequency. The relevant predictions for training on a high precision task (a "difficult" task in RHT labeling, see Jeter et al., 2009) are not specified in the source papers, and so open to interpretation. We suggest that the most likely prediction is that early improvements, and associated transfer, reflect changes at high levels in the visual hierarchy, while subsequent improvements reflect changes at lower, fully specific, levels. The cascade of learning proposed by newer forms of the RHT (Ahissar & Hochstein, 2004) claims that learning first occurs at high and transferrable levels of the visual hierarchy, and then cascades to lower levels of the visual hierarchy if that level leads to improvements in performance. This framework predicts constant and complete specificity with the exception of an early and constant transfer benefit. The current data were not compatible with these claims: either all training should be specific in as much as the task is high precision and so must be learned at a low, nontransferable, level of the visual hierarchy; or some small early amount is transferred, but all subsequent training should have no effect on the amount of transfer. The RHT provides no explanation why an early transfer should be reversed – if performance of the new task based on higher-level learning leads to better performance, then learning of the switch task should begin with the transferrable performance as the starting point of subsequent learning.

We suggest instead that the dynamic properties of specificity in perceptual learning are better understood as the learned optimization of the selection or weighting of sensory inputs to the task (Dosher & Lu, 1998; Petrov et al., 2005). During training, a first-approximation to the optimum connections (weights) that selectively enhances the channels near the signal stimuli and downweights the task-irrelevant channels are learned first, and a more specific weight optimization that more narrowly focuses on the signal stimuli is refined and solidified as learning continues. This possible interpretation is generally consistent with an Augmented Hebbian Reweighting Model (AHRM) of perceptual learning (Petrov et al., 2005). It may also provide an explanation for reported differences in brain activation during different phases of perceptual learning (Yotsumoto, Watanabe, & Sasaki, 2008).

Petrov et al. (2005) trained observers in an orientation discrimination task, alternating fairly extended training in each of two distinct external noise contexts, and found that the AHRM model (for learning in a single location) predicted two aspects of learning also seen in the data: a general improvement in performance over practice, and a cost for switching task environments that persisted over multiple back and forth task changes. Adaptive reweighting of sensory representation inputs to a decision unit over practice changes the weight profile to increasingly focus on relevant spatial frequencies and orientations, while at the same time task-specific weight optimization caused substantial and persistent costs at task switches. The experiment in the current paper differs in the change of angles over tasks, rather than external noise characteristics, and also by training in a new visual location; it also measures only two task-training phases. Even so, the current data pattern is similar to the early phases of the Petrov et al. model predictions and data: Extensive training on a first task should produce general improvements, especially early in training, offset by increasingly specific weight optimization that will increase the switch cost at the change of tasks. The principles of general improvement increasingly offset

---

[6] These switchback results may depend upon ongoing general learning and the distance between the orientations trained in the two tasks.

by switch costs associated with highly specific optimization after extensive training are consistent with the current data set.

This incremental reweighting framework also provides a possible explanation for the current "switchback" condition measurements, which are consistent with the observations of Petrov et al. (2005) for the earliest switch costs, where general improvements are offset by switch costs to yield a "switchback" at about the last level on the previous task. This interpretation is only distinguished in the Petrov et al. data by the successive, multiple, switch design – a little used but powerful design for distinguishing between independent learning and push–pull optimization of alternate task performance.

According to the AHRM model, switch costs are specific to cases where the two tasks have significant differences between shared parts of the optimized weight structures. In the case where the optimum weight structures for the two tasks are closely similar and generally consistent with one another, cooperative learning may occur in addition – or instead. This framework suggests that the finding of increased training leading to increased switch costs for the second task likely reflects the switch of angles between the two tasks. A location switch without a switch of angle might very well show less specificity with extended training. Development of a fully implemented AHRM model that generates quantitative predictions for transfer across retinal locations, a multi-location AHRM, is a substantial independent project.

In sum, we favor the incremental reweighting framework's account of the current data on the extent of training and the effects on transfer and specificity. Our results and conclusion will benefit by further testing either in related between-group designs on different task combinations, or in related task paradigms. Together with consideration of interesting recent reports of location double training on transfer to different locations in perceptual learning Xiao et al. (2008), this work addresses one of the most important questions in perceptual learning: why and how does learned perceptual expertise transfer?

Our study documented the effect of different amounts of transfer by comparing learning in different groups of observers. An alternative approach (suggested by a reviewer) might have been to assess transfer on a second task after different stages of perceptual learning in a primary task. The alternative within-observer design is complicated by two factors. First, the work of Xiao et al. (2008) showed that alternate training of locations may in some circumstances "promote" transfer from one location to another. The effects of double training are not fully understood yet; the comparison of transfer in groups receiving different amounts of training avoids them. Second, the periodic assessment of performance on a transfer task within an observer is a variant of the Petrov et al. (2005, 2006) paradigm in which observers alternate between two learning or task contexts, perhaps with only a single block of practice in the transfer "assessment" phases interspersed with longer training phases on the main task. The Petrov et al. study, described above, found general learning as well as optimization for each task/context and consequent switch costs following every task alternation. These results document that the transfer assessment phases of a within-observer design would alter the system it is trying to measure – a form of Heisenberg principle in perceptual learning. Interpretation of such data, as suggested above, would require a quantitative model – a new elaboration of the Petrov et al. (2005) Augmented Hebbian Reweighting Model (AHRM) (or a contender model) – for transfer across different retinal locations to provide a system within to assess the potentially complex interactions of primary and transfer training. The substantive theoretical elaboration and experimental testing of such a model remain for future investigations. Together with critical empirical tests, these model developments aim to contribute to a broad theoretical account of specificity and transfer in perceptual learning.

References

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

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*(10), 457–464.

Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perceptual-learning in parafoveal vision. *Vision Research, 35*(12), 1679–1690.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Censor, N., & Sagi, D. (2009). Global resistance to local perceptual adaptation in texture discrimination. *Vision Research, 49*, 2550–2556.

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.

Dill, M. (2002). Specificity versus invariance of perceptual learning: The example of position. In M. Fahle & T. Poggio (Eds.), *Perceptual learning* (pp. 219–229). MIT Press.

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

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*(19), 3197–3221.

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 of the United States of America, 102*(14), 5286–5290.

Dosher, B. A., & Lu, Z. L. (2006). Level and mechanisms of perceptual learning: Learning first-order luminance and second-order texture objects. *Vision Research, 46*(12), 1996–2007.

Dosher, B. A., & 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. A., & Lu, Z. L. (2009). Hebbian reweighting on stable representations in perceptual learning. *Learning and Perception, 1*, 37–58.

Fahle, M., & Poggio, T. (2002). *Perceptual learning.* Cambridge: The MIT Press.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision, 2*(2), 190–203.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination – Specificity for orientation and spatial-frequency. *Vision Research, 21*(7), 1149–1158.

Fisher, R. A. (1932). *Statistical methods for research workers (4th ed.).* Edinburgh, London: Oliver and Boyd.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*(5), 681–697.

Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*(10), 1055–1056.

Huang, C. B., Zhou, Y. F., & Lu, Z. L. (2008). Broad bandwith of perceptual learning in the visual system of adults with anisometropic amplyopia. *Proceedings of the National Academy of Sciences of the United States of America, 105*(10), 4068–4073.

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

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture-discrimination – Evidence for primary visual-cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*(11), 4966–4970.

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), 467+.

Li, R. J., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience, 12*(5), 549–551.

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

Lu, Z. L., Chu, W., Dosher, B. A., & Lee, S. (2005). Independent perceptual learning in monocular and binocular motion systems. *Proceedings of the National Academy of Sciences of the United States of America, 102*(15), 5624–5629.

Lu, Z. L., & Dosher, B. (2009). Mechanisms of perceptual learning. *Learning & Perception, 1*(1), 19–36.

Lu, Z. L., Liu, J. J., & Dosher, B. A. (2010). Modeling mechanisms of perceptual learning with augmented Hebbian re-weighting. *Vision Research, 50*(4), 375–390.

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

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

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual-learning in visual hyperacuity. *Science, 256*(5059), 1018–1021.

Polat, U. (2010). Making perceptual learning practical to improve visual functions. *Vision Research, 49*, 2566–2573.

Ramachandran, V., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception, 2*(3), 371–376.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual-learning in identifying the oblique orientation – Retinotopy, orientation specificity and monocularity. *Journal of Physiology – London, 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.

The Mathworks, I. (1999). *MATLAB 5.2*. MA: Natick.

Webb, B. S., Roach, N. W., & McGraw, P. V. (2007). Perceptual learning in the absence of task or stimulus specificity. *PLoS ONE, 2*(12), e1323. doi:10.1371/journal.pone.0001323.

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*, 1922–1926.

Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron, 57*(6), 827–833.

Yu, C., Klein, S., & Levi, D. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision, 4*(3), 169–182.

Zhang, T., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Decoupling location specificity from perceptual learning of orientation discrimination. *Vision Research, 50*, 368–374.