The relationship between initial threshold, learning, and generalization in perceptual learning

Gábor Lengyel

Department of Cognitive Science,
Central European University, Budapest, Hungary   ✉

József Fiser

Department of Cognitive Science,
Central European University, Budapest, Hungary   ✉

We investigated the origin of two previously reported general rules of perceptual learning. First, the initial discrimination thresholds and the amount of learning were found to be related through a Weber-like law. Second, increased training length negatively influenced the observer's ability to generalize the obtained knowledge to a new context. Using a five-day training protocol, separate groups of observers were trained to perform discrimination around two different reference values of either contrast (73% and 30%) or orientation (25° and 0°). In line with previous research, we found a Weber-like law between initial performance and the amount of learning, regardless of whether the tested attribute was contrast or orientation. However, we also showed that this relationship directly reflected observers' perceptual scaling function relating physical intensities to perceptual magnitudes, suggesting that participants learned equally on their internal perceptual space in all conditions. In addition, we found that with the typical five-day training period, the extent of generalization was proportional to the amount of learning, seemingly contradicting the previously reported diminishing generalization with practice. This result suggests that the negative link between generalization and the length of training found in earlier studies might have been due to overfitting after longer training and not directly due to the amount of learning per se.

Introduction

In the last decades, numerous factors were identified that influence observers' ability to improve their performance in low-level perceptual tasks after extensive practice, a process termed perceptual learning. Among these factors are feedback (Aberg & Herzog, 2012; Herzog & Manfred, 1997; Petrov, Dosher, & Lu, 2006; Seitz & Watanabe, 2003), experimental design (Adini, Wilkonsky, Haspel, Tsodyks, & Sagi, 2004; Kuai et al., 2005; Yu, Klein, & Levi, 2004), the nature of the contextual elements around the target (Adini, Sagi, & Tsodyks, 2002; Manassi, Sayim, & Herzog, 2012), or more broadly, the structure and the variability of the stimuli and the task (Y. Cohen, Daikhin, & Ahissar, 2013; Hussain, Bennett, & Sekuler, 2012; Kuai et al., 2005). More recently, the generalization of learning in perceptual tasks also came under investigation, and once again, researchers identified a great number of factors that determine the extent of generalization. Among others, task difficulty (Ahissar, Merav, & Shaul, 1997), precision (Jeter, Dosher, Petrov, & Lu, 2009), stimulus variability (Hussain et al., 2012), training length (Ahissar et al., 1997; Jeter, Dosher, Liu, & Lu, 2010), additional tasks and stimuli (Hung & Seitz, 2014; Wang, Zhang, Klein, Levi, & Yu, 2014; Xiao et al., 2008; Zhang et al., 2010), and statistical structure of the task and stimuli (Y. Cohen et al., 2013) have an effect on the level of generalization. Although these studies broadened our understanding of the underlying processes of perceptual learning, only few of them can provide support for general rules that could predict perceptual learning performance under different conditions (e.g., Ahissar et al., 1997; Astle, Li, Webb, Levi, & McGraw, 2013; Hussain et al., 2012; Jeter et al., 2010). The present study focuses on two previously investigated more universal rules that were suggested to predict performance in perceptual learning paradigms in general: the link between initial performance and the magnitude of perceptual learning (Astle et al., 2013), and the connection between the amount of learning and the extent of generalization (Hussain et al., 2012, Jeter et al., 2010).

Comparing initial performance and the amount of learning

Several studies reported that the amount of learning in perceptual tasks (as defined by the improvement in

performance from the first day to the last one) can be predicted from the initial performance (Aberg & Herzog, 2009; Astle et al., 2013; Fahle, 1997; Fahle & Henke-Fahle, 1996; Polat et al., 2012, Yehezkel, Sterkin, Lev, Levi, & Polat, 2016). The earlier examples of these studies used one-interval 2-AFC hyperacuity tasks (Vernier, curvature, and orientation discrimination tasks). These studies found that the better observers' initial performance was the smaller they improved on the task (Fahle, 1997; Fahle & Henke-Fahle, 1996).

However, a more recent study by Astle and colleagues (2013) investigated this relationship in more depth and argued for a specific, Weber-like relationship (Fechner, 1860; Ross & Murray, 1996) between the initial performance and the magnitude of learning. In their study, monocular Vernier acuity was measured at various eccentricities in a one-interval 2-AFC task after observers were trained at both 5° and 15° off the central fixation. The authors found that the initial discrimination thresholds, on average, were higher at 15° eccentricity than at 5°. Critically, the amount of improvement on the Vernier acuity task (measured as a difference of the first and the final day's Vernier discrimination threshold in arcsec) was proportional to the initial discrimination thresholds (Astle et al., 2013). In addition, when they equated the observers' initial thresholds at the various eccentricities in the acuity task by spatially scaling the Vernier lines or by visual crowding, the magnitude of learning became equal at the different eccentricities. Thus, regardless of what constraint limited the initial discrimination thresholds prior to training (retinal location, stimulus size, or crowding), the amount of absolute learning seemed to be proportional to the initial threshold level. To further specify this claim, Astle et al. (2013) expressed the relative learning as the observers' ratio of the first and the final day's thresholds (measured in Vernier discrimination threshold in arcsec divided by the line length also in arcsec) and showed that this relative learning did not correlate with the initial thresholds, but it was constant across different levels of initial thresholds. Because this pattern is captured by Weber's law, Astle and colleagues posited that "…perceptual learning also obeys a similar Weber-like law…" and that "…the finding that improvements in normal subjects are tied to their initial threshold in a lawful way, analogous to Weber's law, suggests that the same factors that impose limits on a visual threshold also constrain the amount an organism can learn on a visual task." (Astle et al., 2013, pp. 4 and 7). Astle and colleagues' results (a Weber-like law for absolute learning leading to no correlation in terms of relative learning) are in contrast with those of earlier studies that used the same measure of relative learning, but did report a positive correlation between the relative

learning and initial performance in Vernier (Fahle & Henke-Fahle, 1996) and bisection acuity tasks (Aberg & Herzog, 2009). The positive correlations found in those studies means that the amount of absolute learning measured in those experiments was a progressively increasing fraction of the initial discrimination thresholds, implying a power-like law (Stevens, 1957) rather than a Weber-like law.

The discrepancy between the results of the above studies can be tracked back to the issue of whether the relationship between learning and initial threshold is influenced by something else beyond the observers' perceptual scaling function. In psychophysics, the observer's perceptual scaling function represents how physical stimulus intensities are related to perceived magnitudes. Assuming that the discrimination threshold is limited by constant and independent Gaussian noise in accordance with signal-detection theory in its most basic form (Green & Swets, 1966), the perceptual scaling function can be estimated by measuring the observer's discrimination thresholds at different physical stimulus intensities. As the discrimination thresholds represent the lowest increment in the stimulus intensity that the observer can still perceive (at a certain performance level), the scaling function approximates how the observer maps the physical stimulus onto her internal perceptual space (see Figure 1, top). In this paper, we argue that the proportional Weber-like relationship between initial discrimination thresholds and the amount of learning (Figure 1, bottom) emerges in perceptual learning tasks when (a) observers improve by the same amount at different region in their internal perceptual intensity space, and (b) the perceptual scaling function between the perceptual and physical spaces does not change during learning. In this case, the amount of learning will depend only on the same perceptual scaling function of the observer that also determines her initial discrimination threshold prior to learning. Consequently, there will be a proportional relationship between initial threshold and the amount of learning. In contrast, power-like law (or any *not* proportional functional relationship) between initial discrimination thresholds and learning would emerge only when, in addition to the perceptual scaling function, either a change in the perceptual scaling due to learning or some other additional learning-specific factors affect perceptual learning.

Figure 1 shows two simple examples demonstrating the argument above. In the top plots on the left, the hypothetical observer has a Weber-like perceptual scaling function that transforms physical intensity to perceived magnitude. In the plots on the right, the hypothetical observer has a power-like perceptual scaling function. In both cases, the same amounts of improvement in the perceptual intensity space at different base-intensities (colored ranges on *y* axis) will
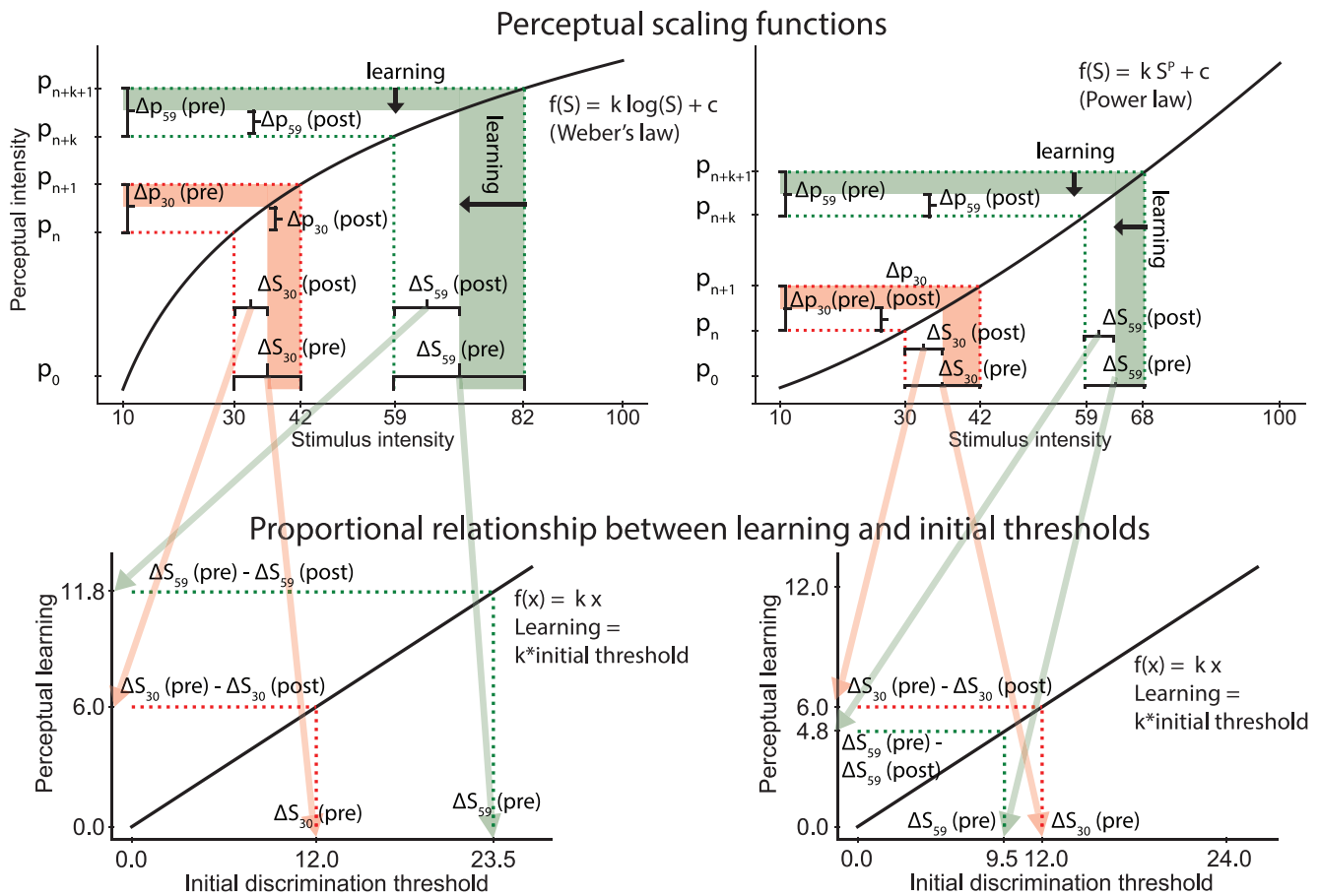
Figure 1. The relationship between initial discrimination thresholds and the amount of learning, and how this relationship is related to observers' perceptual scaling function linking physical and perceived intensities. Top: Two typical perceptual scaling functions found in human perception: the Weber's law (Left) and the Power law (Right). Physical intensities on the *x* axis show a hypothetical scale of a visual attribute from 10 to 100, while the perceptual intensities on the *y* axis scale from the absolute threshold (P_0). The scale on the *y* axis depends on the function that maps the physical magnitudes onto the perceptual intensities. Two initial discrimination threshold levels at two base-intensities are shown, at 30, ΔS_{30}(pre), large black brackets between the red dotted lines and at 59, ΔS_{59}(pre), large black brackets between the green dotted lines. These initial discrimination thresholds (i.e., *just noticeable differences*, JNDs) reveal the smallest step sizes on the stimulus intensity space (*x* axis) that have a corresponding one unit change on the observers' perceptual space (*y* axis), Δp_{30}(pre) and Δp_{59}(pre) the smallest perceivable changes at the measured base-intensities. Assuming the same amounts of learning measured by the perceptual sensitivity at the two base-intensities, these unit sizes decrease with the same amount at the two base-intensities on the perceptual intensity space represented by the colored ranges on the *y* axis. This equal amount of improvement in the perceptual space will be transformed through the perceptual scaling function back into changes in the stimulus intensity (colored changes on the *x* axis) which therefore, will be proportional to the initial thresholds. Hence, the amounts of perceptual learning at different base-intensities {ΔS_{30} (pre) − ΔS_{30} (post)} and {ΔS_{59} (pre) − ΔS_{59} (post)} will follow the same perceptual scaling function that determined the initial discrimination thresholds {ΔS_{30}(pre) and ΔS_{59}(pre)} prior to learning. Consequently, proportional relationship between the initial discrimination thresholds and the amount of learning at the two base-intensities emerges: $\frac{\Delta S_{30}(PRE)}{\Delta S_{59}(PRE)} = \frac{\Delta S_{30}(PRE) - \Delta S_{30}(POST)}{\Delta S_{59}(PRE) - \Delta S_{59}(POST)}$ Bottom: The proportional relationship between initial thresholds (IT) and perceptual learning (PL). Initial discrimination thresholds, ΔS(pre) are shown on the *x* axis, while the amount of learning, ΔS(pre) − ΔS(post) on the *y* axis. The dotted red and green lines represent the corresponding initial discrimination threshold levels and the amount of learning at 30 and 59 stimulus intensities derived from the top panels. The green and red arrows show the relationship between the top and the bottom figures for the two stimulus intensities. Regardless of the exact perceptual scaling function (progressively increasing, power-like or progressively decreasing Weber-like function) the relationship between learning and initial thresholds remains proportional: $PL = k \times IT$, with *k* as a scaling constant.

lead to different amounts of improvement in the stimulus intensity space (colored ranges on x axis) depending only on the shape of the perceptual scaling function linking physical and perceptual intensities. Therefore, both the initial thresholds (pre) and the amounts of learning (pre − post) follow the same function, the observers' perceptual scaling function. This condition will automatically lead to changes in the amount of learning that is proportional to the initial thresholds. This theory is only true if (a) the shape of the perceptual scaling function does not change during learning, and (b) the same amounts of improvement occur on the internal perceptual space at the different base-intensities (such as the red and green shaded areas on the y axis). However, this linear proportionality vanishes if either the observers' functional mapping from physical to perceptual intensities is modulated during learning or the amounts of improvement on the internal perceptual space are different at different base-intensities.

Using the above observations, the difference between the findings of Astle et al. (2013) and Aberg and Herzog (2009) and Fahle and Henke-Fahle (1996) can be captured as follows. Astle et al.'s result can be explained by assuming that (a) observers improve by the same amount at different base-intensities in their internal perceptual space, and (b) their perceptual scaling function does not change during perceptual learning. In this case, the amount of learning depends only on the observers' perceptual scaling function, without assuming any learning-specific extra Weber-like process they posit in their paper. In contrast, assuming a change in the scaling function during learning and/or different amounts of learning at different base-intensities in the internal perceptual space would distort the Weber-like proportionality between initial threshold and learning, confirming Aberg and Herzog's and Fahle and Henke-Fahle's results. In this case, the amount of learning cannot be predicted from the observers' scaling function suggesting that other learning-specific processes are involved during perceptual learning. The first goal of the present study was to investigate which of these two scenarios hold in general during perceptual learning.

The relationship between the amount of learning and generalization

Traditionally, the specificity of the acquired ability has been a defining hallmark of perceptual learning (Crist, Kapadia, Westheimer, & Gilbert, 1997; Fahle, 1997; Karni & Sagi, 1991; Schoups, Vogels, & Orban, 1995). According to this view, whatever improved ability observers develop after extensive training within the context of low-level visual discrimination tasks, this

new skill remains available only within the close context of the original setup including the stimulus identity and the location of training in the retinal space. However, recent studies finding substantive transfer of learning under various conditions strongly challenge this notion (Ahissar & Hochstein, 2004; Wang et al., 2014; Zhang et al., 2010).

Investigating the relationship between the amount of learning and generalization involves an inherent ambiguity at the conceptual level. Intuitively, generalization and learning should go hand in hand: More learning means more knowledge about the state of the world and hence, more potential for using the newly learned competence in different contexts. However, it is well known in the field of machine learning that too much repetitive learning can result in a representation (an internal model) that is overly specific to the trained features and the circumstances of the training, a phenomenon called overfitting (Hastie, Tibshirani, & Friedman, 2009; Murphy, 2012). In perceptual learning, learning can be defined as the improvement in task performance in a context-specific manner (in the trained condition), while generalization is the improvement in task performance in a context-free manner (in an untrained condition). Overfitting is related to the difference between the two. Excessive training in perceptual learning could cause overfitting, which could lead to a little extra learning, but it also substantially decreases generalization. Indeed, several behavioral studies in the domain of perceptual learning confirmed this conjecture (Hussain et al., 2012; Jeter et al., 2010).

Thus, the relationship between perceptual learning and generalization can depend intricately on two separate components: Whereas the specific features of the learning process, such as the selected task or the training stimuli, could lead to more specific or more generalizable knowledge, overtraining itself can shift performance from flexible to specific. Since overfitting is a general rule in computational learning theories, we were interested in exploring the first component, whether more perceptual learning produces more generalizable knowledge before the effect of overfitting emerges. If more training in various perceptual tasks leads to more learning due to an improved internal model incorporating the actual experience in the observer's world model, proportionally more generalization is predicted before overfitting occurs. However, if more training results in more learning due to focusing only on specific features of the task/stimuli without viable integration of this knowledge to other aspects of the observer's internal model, learning is expected to be proportionally more specific to the features of the training examples even before overtraining happens. Previous studies modulated the extent of training (Hussain et al., 2012; Jeter et al., 2010) which, although

influenced the amount of learning, also increased the amount of training data from the same kind rather than providing more new information with the training data which increases the chance for overfitting (Hastie et al., 2013; Murphy, 2012). We used a five-day long fixed length training protocol to control for the effect of overfitting and measured the individual differences in the amount of learning and generalization in two widely used perceptual learning paradigms (contrast and orientation discrimination tasks). This set up allowed us to pursue the second goal of the present study, to determine whether the extent of generalization is proportional to the amount of learning.

Overview of the present study

In three experiments, we measured contrast and orientation discrimination thresholds and the amount of learning at two different stimulus intensities (at 73% and 30% contrast, and at 25° and 0° orientation in separate temporal 2-AFC discrimination tasks) and found that the amount of perceptual learning was proportional to the initial performance. Furthermore, we showed that this specific relationship between initial performance and learning mainly reflected the observers' internal perceptual scaling function which transforms physical magnitudes to perceptual intensities. Our results also revealed a positive link between the amount of learning and generalization: More learning led to proportionally more generalization. We interpreted the relationship between these results and earlier reports in the light of differences in methodological and conceptual characteristics of perceptual learning paradigms.

Methods

Observers

One hundred and twenty naive observers gave informed consent prior to participation in the experiment. Nineteen observers took part in Experiment 1, the within-subject contrast discrimination experiment. In Experiment 2, 25 observers completed the 30% reference condition and another 24 observers the 73% reference contrast condition. In the orientation discrimination experiments (Experiment 3), 15 and 15 observers participated in the 0° and 25° reference conditions, and another 11 and 11 observers completed the 15° and the 45° reference orientation conditions, respectively. None of the observers had any previous experience with a psychophysical experiment. All participants had normal or corrected-to-normal vision.

The experimental protocols were approved by the Ethics Committee for Hungarian Psychological Research.

Apparatus and stimuli

We used MATLAB (MathWorks, Natick, MA) Psychtoolbox 3 (Brainard, 1997; Pelli, 1997) to generate the stimuli on a 21-in Samsung Syncmaster 1100 DF color monitor (1024 × 768, 85 Hz frame rate, 0.2 mm pixel pitch). The mean luminance was 60 cd/m^2. The monitor was calibrated, and the luminance was linearized by X-Rite i1Profiler device and software. The participants viewed the stimuli binocularly at the fovea in a dimly lit room. In both paradigms, the stimuli were Gabor patches defined by Gaussian enveloped sinusoidal gratings with (spatial frequency of 6 cycles/° (*SD*: 0.17°), contrast 0.47 in the orientation discrimination task, orientation 36° in the contrast discrimination tasks, and phase randomized for every stimulus presentation in the orientation discrimination task). The Gabor patches were presented on a background at mean luminance. The stimuli were viewed from 2 m through a circular aperture (diameter 17°) of a black piece of cardboard that covered the entire monitor screen. The whole cardboard and the viewing area in front of the observer were further covered by a black curtain with a circular aperture (diameter 17°). This setup was used to prevent observers from using the edges of the display in the orientation discrimination task.

Procedure

Investigating initial thresholds and the amount of learning

We conducted perceptual learning experiments using two attributes, contrast and orientation, and we measured discrimination thresholds from a reference value. To test whether perceptual learning was proportional to the initial performance due to the internal perceptual scaling of the participants, we used two experimental conditions in all experiments. In the two conditions, observers were trained with two different stimulus intensities that were known to elicit different initial discrimination threshold levels according to previous studies measuring the perceptual scaling functions of the observers between physical and perceptual magnitudes. In Experiments 1 and 2 the two conditions were distinguished by the reference contrast (30% vs. 73%) at which the observers were trained. Based on previous studies (Burton, 1981; Legge, 1981), we expected significantly higher initial discrimination thresholds at 73% contrast. In Experiment 3, observers
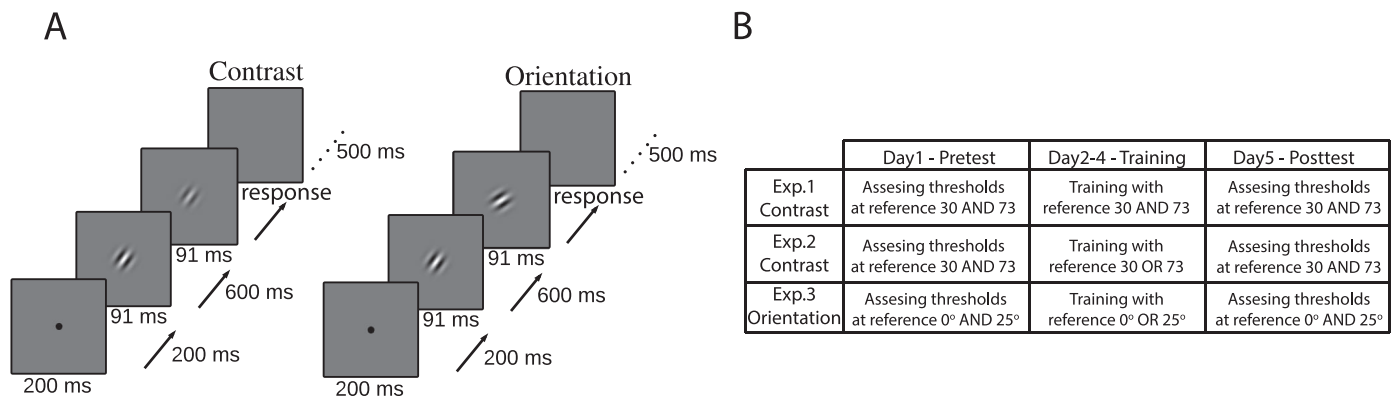
A

B

	Day1 - Pretest	Day2-4 - Training	Day5 - Posttest
Exp.1 Contrast	Assesing thresholds at reference 30 AND 73	Training with reference 30 AND 73	Assesing thresholds at reference 30 AND 73
Exp.2 Contrast	Assesing thresholds at reference 30 AND 73	Training with reference 30 OR 73	Assesing thresholds at reference 30 AND 73
Exp.3 Orientation	Assesing thresholds at reference 0° AND 25°	Training with reference 0° OR 25°	Assesing thresholds at reference 0° AND 25°

Figure 2. (A) Contrast and orientation discrimination tasks. (B) Training protocol.

were trained at reference orientation of 0° versus 25°. Again, since earlier studies reported the lowest discrimination threshold at the cardinal orientations (Mansfield, 1974; Mikellidou, Cicchini, Thompson, & Burr, 2015; Orban, Vandenbussche, & Vogels, 1984, Regan & Price, 1986), we expected higher initial discrimination thresholds at 25°. Once the initial discrimination thresholds from the reference values were measured, we assessed the amount of perceptual improvement in each of the conditions and checked whether they showed proportionally more learning in the conditions with higher versus lower initial discrimination threshold levels.

There is a trade-off in benefits when using within- versus between-subject designs in perceptual learning tasks. On the one hand, a related samples statistical analysis in a within-subject design is more sensitive, and therefore, it can reveal a relationship between initial performance and learning even if the individual differences in perceptual performances are large. On the other hand, a within-subject design is potentially prone to uncontrolled generalization between the conditions, which can bias the comparison between the low and high initial performance conditions. To control for this problem, participants in Experiment 1 trained with both reference contrast conditions, 30% and 73% (within-subject design, Figure 2B) similarly to Astle et al.'s (2013) study, in which participants trained at both 5° and 15° eccentricities. Meanwhile, in Experiments 2 and 3, two separate groups of observers were trained with either 30% or 73% reference contrasts in the contrast discrimination task, and with either 0° or 25° in the orientation discrimination tasks (between-subject designs, Figure 2B).

Investigating generalization and the amount of learning

Generalization was quantified by measuring discrimination thresholds at an untrained reference contrast or orientation after finishing the training sessions. In Experiment 1 after training with reference contrast at 30% and 73%, generalization was assessed by measuring discrimination threshold at the untrained 47% contrast. In Experiment 2 for the group that practiced with reference contrast at 30%, generalization was tested at both 47% and 73% contrast levels; and for the group that trained with reference contrast 73%, the transfer of learning was tested at contrast of 30% and 47%. In the orientation discrimination task, generalization in the group that trained at 0% reference orientation was measured at 25°, and in the group that practiced with reference orientation 25° it was assessed at 0°. We tested whether more learning caused proportionally more generalization by assessing the within-condition correlation between individual differences in learning and in generalization. Due to very small intersubject variability in perceptual performances at 0° reference orientation, two additional groups of participants completed the very same experiment, but one group trained with reference orientation 15° and generalization was measured at 45°, and the other group trained with reference orientation 45° and generalization was assessed at 15°. In these groups we had sufficiently large intersubject variability to test our question about generalization (see Supplementary materials and Supplementary Figure S3).

General procedure

Contrast and orientation discrimination thresholds were measured with a temporal two-alternative forced choice (2-AFC), 1-up-3-down staircase procedure. In each trial, a fixation point was first flashed for 200 ms and disappeared 200 ms before the onset of the first stimulus interval. Next, the reference (contrast or orientation) and test patch were presented after each other for 91 ms each in a random order. The reference and the test patch were separated by a 600 ms interstimulus interval (Figure 2A).

In all experiments, observers trained for five consecutive days, completing one session per day (Figure 2B). In each trial, the observer had to judge

whether the stimulus has a more clockwise orientation (in the orientation discrimination task) or a higher contrast (in the contrast discrimination task) in the first or the second stimulus interval. Observers responded by pressing "1" or "2" keys on the keyboard. In all tasks, there was an auditory feedback marking incorrect responses.

The staircase during the experiments followed the 3-down-1-up rule with a step size of 0.05 log units, which converged to 79.4% correct responses (Levitt, 1971). The initial difference values between the reference and the test for the very first staircase were $\Delta 8\%$, $\Delta 12\%$ for reference contrasts 30% and 73%, $\Delta 8°$ and $\Delta 14°$ for reference orientation 0° and 25°, and $\Delta 8°$ and $\Delta 18°$ for reference orientation 15° and 45°, respectively. The initial differences were determined based on the mean initial discrimination thresholds of the observers in our pilot perceptual learning experiments using the same procedure to approximate contrast and orientation discrimination thresholds. After completing the first staircase, the initial values for the following staircases were adjusted separately for each observer by taking the observer's average performance in the previous staircase in the same condition and multiplying it by two. Each staircase contained four practice and six experimental reversals. The observer's threshold was defined as the geometric mean of the experimental reversals. Observers completed 5-5 staircase blocks in each reference value conditions in the pre-and posttest sessions and 10 staircase blocks with the practiced reference value during each training session. Previous results using simulations suggested that the adaptive method described above should reveal observers' thresholds at 79.4% performance level quite accurately (García-Pérez 1998). However, in those simulations attentional lapse rates were assumed to be zero, and estimating discrimination thresholds based on the stimulus strengths at reversal points could be confounded by attentional lapses (Solomon & Tyler, 2017). Although theoretical work and simulations showed that the 3-down-1-up staircase is robust to the initial attentional lapses (Karmali, Chaudhuri, Yi, & Merfeld, 2016), lapses are not necessarily limited to the initial trials in novice observers. In order to confirm that the measured decrease in thresholds after practice using the 3-down-1-up staircase method was not just due to the decrease in attentional lapses of our participants, we estimated the lapse rates and the thresholds for each observer by fitting psychometric curves to their performance at pre- and posttests (see the detailed methods in Supplementary materials). We found that the thresholds decreased significantly after the training in all experimental conditions (see Supplementary Figure S1) even when we controlled for the decrease in lapse rates (see Supplementary Figure S2). Furthermore, the decrease in thresholds due to learning estimated by the best-fitting threshold parameter of the participants' psychometric

curves positively correlated with the estimated decrease in thresholds using the adaptive staircase method (see Supplementary Figure S1–S2). This suggests that perceptual learning measured by the thresholds at pre- and posttests using the staircase method reveals perceptual and not just attentional improvement.

Analysis

Exclusion criteria

We excluded outlier participants from the analysis if their performance (in initial thresholds or learning) was more than 2 *SD* away from the group average. Using this criterion, we excluded two subjects from Experiment 1 because one of them had large negative learning in the reference contrast 30%, and the other one had large negative learning in the reference contrast 70% conditions. We excluded one-one subjects from each of the conditions in Experiment 2 for the same reason: Both participants showed large negative learning. There were no outliers in the orientation discrimination task; thus, we did not exclude anyone from the analysis in Experiment 3.

Assessing learning

To measure the amount of perceptual learning we used three types of learning scores.

a. Absolute learning computed as $PL^{abs} = PRE - POST(thresholds)$.
b. Relative learning computed as $PL^{rel} = \frac{PRE}{POST}(thresholds)$.
c. Predicted learning computed as (see Results and discussion), $PL^{predicted} = \frac{PRE_{@LowStimIntensity}}{PRE_{@HighStimIntensity}} \times PL^{abs}$.

Assessing generalization

The amount of generalization was assessed with two metrics: by computing *absolute generalization* as Pre − Post thresholds at the untrained reference values, and by computing *relative generalization* as the absolute generalization divided by the amount of learning (i.e., generalization/learning).

Statistical analyses

Comparing group means: In our analysis, we needed to evaluate the probability of no difference between two groups' scores, and the probability of certain scores not being different from zero. However, frequentist hypothesis testing cannot confirm the null hypothesis due to its design (Morey & Rouder, 2011; Streiner, 2003).Therefore, we ran independent or paired samples *t* tests and also nonoverlapping hypotheses (NOH) Bayes factor (BF) analysis for independent or related

samples (Morey & Rouder, 2011) to compare the different conditions in the experiments. We computed the Nonoverlapping hypotheses (NOH) Bayes factors (BF; Morey & Rouder, 2011; Rouder, Speckman, Dongchu, Morey, & Geoffrey, 2009) to obtain the level of confidence in concluding no difference between certain learning scores (see Results for the specific comparisons). The NOH BF represents the probability of "there is no or negligible difference between the conditions" divided by the probability of "there is difference between the conditions." Therefore, BFs larger than one indicates how many times more probable the no or negligible difference than the existence of a difference between the conditions is. In the NOH BF analysis, the null hypothesis states that the effect size is within the range of −0.2 and 0.2, whereas the alternative hypothesis is that the effect size is outside that range. The range of the null hypothesis was chosen following the guidelines of J. Cohen (1988) and Morey & Rouder (2011) that below 0.2 the effect is negligible. We used a scaling factor equal to one in the scaled Cauchy prior.

Analyzing the variability within conditions: Intersubject variability was analyzed with Pearson and partial correlation. We applied partial correlation between the amount of learning and the extent of generalization while controlling for the initial threshold levels. The partial correlation coefficient reveals the correlation between the residuals of the linear regressions predicting separately generalization and learning from initial thresholds. If the deviations (residuals) from the predicted generalization and from the predicted learning (using the initial discrimination thresholds as predictor in both cases) correlate, it also indicates a relationship between generalization and learning alone without the influence of the initial thresholds. The partial correlation coefficient between X (independent variable) and Y (dependent variable) while controlling for Z (dependent variable) and the standardized regression coefficients of X in a multiple linear regression predicting Y with both X and Z as predictors gives the same amount of information and *p* values. Therefore, computing partial correlation between learning and generalization while controlling for initial threshold levels is equivalent to using multiple linear regression to predict the extent of generalization using the initial threshold levels and the learning scores as independent variables.

Results and discussion

Initial performance and learning

We confirmed that the chosen reference values, indeed, led to groups with higher initial discrimination thresholds at high reference values (73% in the contrast and 25° in the orientation discrimination tasks) than at low reference values (30% contrast and 0° orientation). Specifically, we found significant differences between initial discrimination threshold levels in all experiments: in Experiment 1 ($t_{16} = 7.847$, $p < 0.001$, $d = 1.889$), in Experiment 2 ($t_{45} = 5.852$, $p < 0.001$, $d = 1.664$) and in Experiment 3 ($t_{28} = 6.718$, $p < 0.001$, $d = 2.539$; Figure 3, subpanels A in all panels). This finding means that observers had larger discrimination thresholds around 73% contrast than around 30%, which is in line with previous findings showing a near logarithmic perceptual scaling function from physical to perceived contrast intensity (Burton, 1981; Legge, 1981). In case of the orientation discrimination task, we also found the expected advantage in the discrimination sensitivity at the cardinal orientation (Mansfield, 1974; Mikellidou et al., 2015; Regan & Price, 1986), that is a larger discrimination threshold around 25° than around 0°.

There was significant perceptual learning in all conditions ($ps < 0.005$), although not every observer improved after the training (Figure 3, B and C subpanels in all panels). Perceptual learning was stronger in conditions with higher initial threshold levels (Experiment 1: $t_{16} = 2.567$, $p = 0.021$, $d = 0.693$; Experiment 2: $t_{45} = 2.126$, $p = 0.039$, $d = 0.631$; Experiment 3: $t_{28} = 4.498$, $p < 0.001$, $d = 1.700$; Figure 3, B subpanels in all panels). The ratio of the initial threshold levels and the ratio of the amount of learning in the two conditions were almost the same in all experiments.

$$\text{Experiment 1:} \quad \frac{IT_{Con30}}{IT_{Con73}} = 0.56 \approx \frac{PL_{Con30}}{PL_{Con73}} = 0.51,$$

$$\text{Experiment 2:} \quad \frac{IT_{Con30}}{IT_{Con73}} = 0.53 \approx \frac{PL_{Con30}}{PL_{Con73}} = 0.49,$$

$$\text{Experiment 3:} \quad \frac{IT_{Ori0}}{IT_{Ori25}} = 0.41 \approx \frac{PL_{Ori0}}{PL_{Ori25}} = 0.35,$$

where *IT* and *PL* represent initial thresholds and perceptual learning, respectively.

Whereas these results suggest that the amount of learning is roughly proportional to the initial threshold levels, in the next section we perform a statistical test of the exact proportional relationship and show that it reflects the observers' perceptual scaling function which links physical intensity to perceptual magnitude.

Measuring learning in contrast discrimination by using a within-subject design

In the first contrast discrimination experiment using within-subject design, we tested within each observer
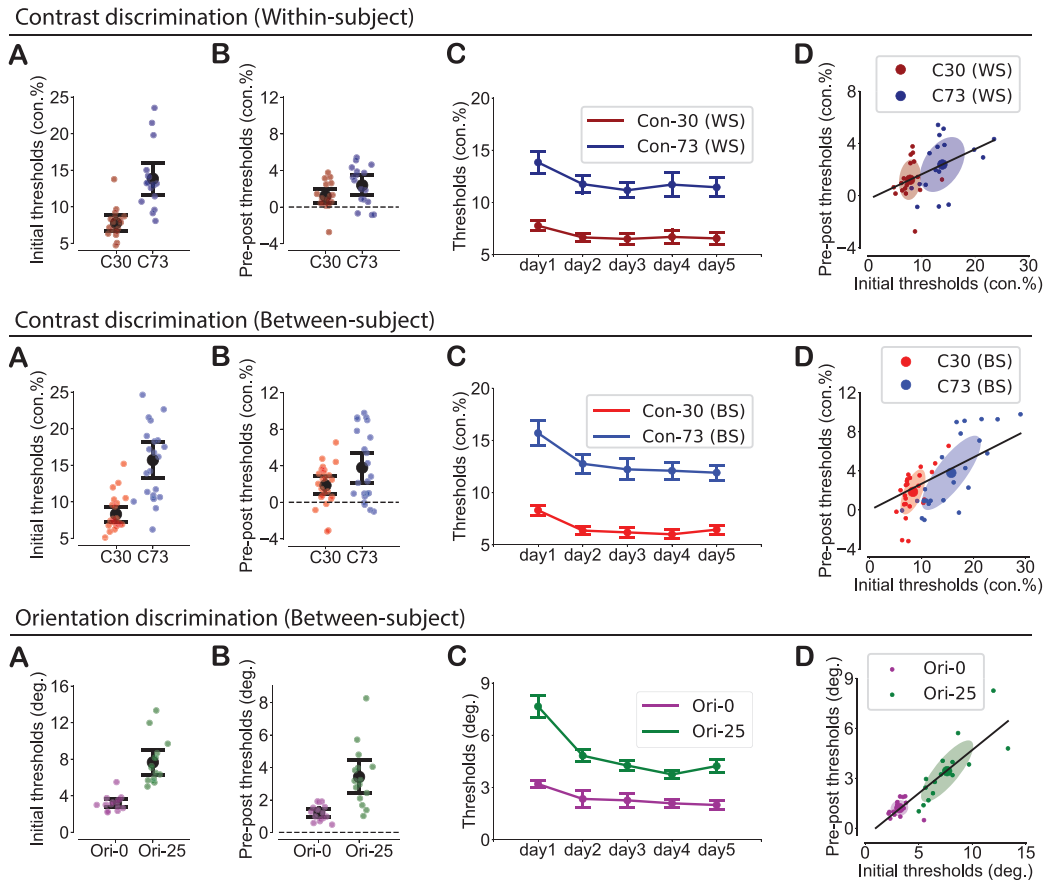
Figure 3. Initial discrimination thresholds and the amount learning. Top panel: contrast discrimination task, within-subject design (WS). Middle panel: contrast discrimination task, between-subject design (BS). Bottom panel: orientation discrimination task, between-subject design. In the contrast experiments red color denotes low (con. 30%) and blue color denotes high reference value conditions (con. 73%). In the orientation experiments purple color denotes low (ori. 0°) and green color denotes high reference value conditions (ori. 25°). In all panels: (A) Initial discrimination thresholds and (B) the amount of absolute learning at the two measured reference values. Error bars represent 95% CI of the mean. (C) Learning curves for the five-day training protocol for the two measured reference values. Error bars show one *SEM*. (D) Learning as a function of initial discrimination thresholds. Error ellipses show one standard deviation, and black lines show linear regression lines fitted to the points from both conditions.

directly whether the amount of observers' learning was proportional to their initial thresholds. The proportionality rule states

$$\frac{IT_{@LowRef}}{IT_{@HighRef}} = \frac{PL_{@LowRef}}{PL_{@HighRef}}, \quad (1)$$

where *@LowRef* and *@HighRef* refer to initial thresholds (*IT*) or perceptual learning (*PL*) assessed at the low (contrast [con.] = 30% and orientation [ori.] = 0°) or high (con. = 73% and ori. = 25°) reference values.

These low and high reference values determined the low and high stimulus base-intensities in our experiments by modulating observers' initial thresholds according to their own perceptual scaling function.

Following Equation 1, we derived the *predicted amount of learning* in the low reference value condition ($PL_{@LowRef}$) by multiplying the left side of Equation 1

with the amount of learning in the high-reference-value condition,

$$\frac{IT_{@LowRef}}{IT_{@HighRef}} \times PL_{@HighRef} = PL_{@LowRef}. \quad (2)$$

For each participant, we computed the predicted amount of learning (left side of Equation 2) at the higher reference value (high base-intensity) and compared it to the *absolute amount of learning* (right side of Equation 2, PRE − POST thresholds) at the low reference value (low base-intensity) within the same observer. If the proportional relationship between the initial thresholds and the amount of learning holds, we expect no difference between the *predicted* and the *absolute learning* scores. Indeed, we found no difference between the two learning scores (Figure 4, top panel A) confirming the proportional relationship between initial thresholds and learning ($t_{16} = 0.216$, $p = 0.832$, $d =$
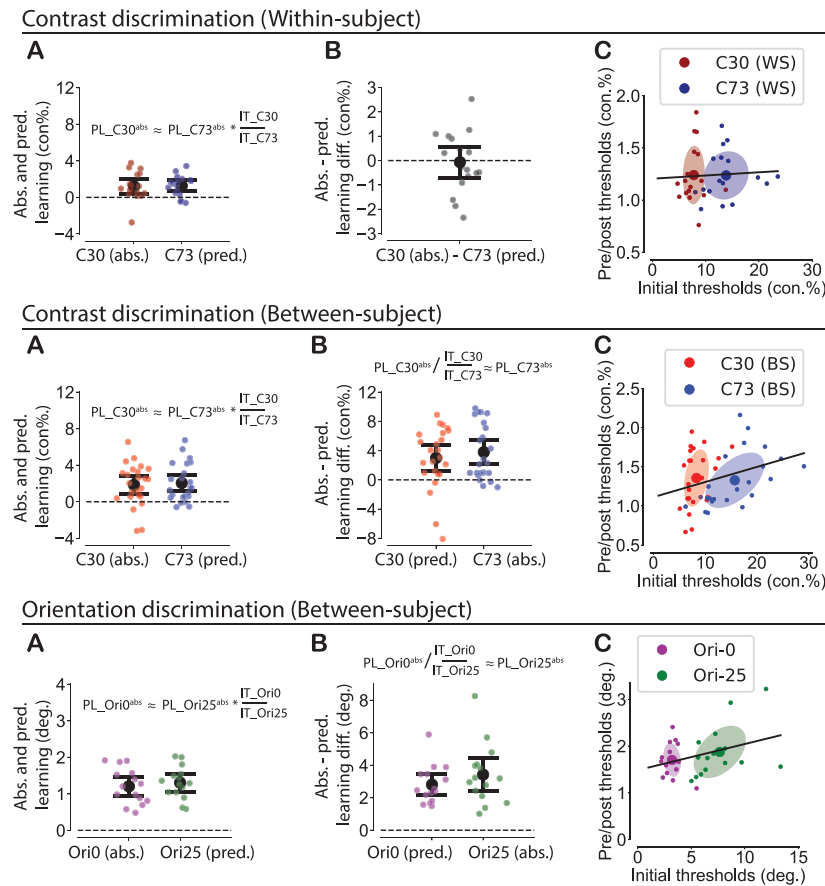
Figure 4. The relationship between initial discrimination thresholds and the amount of learning primarily reflects the observers' scaling function. Top panel: contrast discrimination task, within-subject design (WS). Middle panel: contrast discrimination task, between-subject design (BS). Bottom panel: orientation discrimination task, between-subject design. In all panels: (A) Comparing the absolute learning in the low-reference-value condition to the predicted learning in the high-reference-value condition. (B) Top panel: The difference between the absolute and the predicted amounts of learning at the low and high reference values across subjects. (B) Middle and Bottom panels: Comparing the predicted learning in the low-reference-value condition to the absolute learning in the high-reference-value condition. (A) and (B) Error bars represent 95% CI of the mean, and the equations above the error bars relate absolute to predicted learning in the different conditions derived from Equation 1, capturing the proportional relationship between initial thresholds and learning. (C) Relative learning defined by the ratio of initial discrimination and the posttraining thresholds as a function of the initial threshold levels. Error ellipses show one standard deviation; black lines indicate linear regression lines fitted to the points from both conditions.

0.049, *Bayes Factor favoring no difference* = 10.6). The error bar in Figure 4, top panel B indicate that most of the observers (13/17) deviated less than 1% contrast from the exact proportionality rule as the observers' amount of learning at the two reference values (base-intensities) was almost exactly proportional to their initial threshold levels. This suggests that the individual perceptual scaling functions dominated quite robustly the origin of the proportionality relationship between learning and initial thresholds. The Bayes Factor indicates directly that the "no difference between the learning scores" hypothesis is 10.6 times more probable than "the existence of a difference between the learning scores" (see Methods, Statistical analyses, comparing group means). Therefore, we found strong evidence for the proportional relationship in the data of Experiment 1, and we linked this relationship directly to observers' perceptual scaling functions.

Measuring learning in contrast and orientation discrimination by using a between-subject design

A recurring danger with a within-subject design is the possible confound of cross-training between the conditions, which would allow an alternative explanation to our results in Experiment 1. This calls for an independent confirmation of our findings about proportionality by using a between-subject design. Unfortunately, due to the between-subject design of Experiments 2 and 3 it is not possible to test directly the proportional relationship between learning and the initial thresholds within subjects because separate

groups of observers were trained at the two reference values. However, since the initial thresholds were assessed at both reference values in each group, one could use Equation 1 to calculate the *predicted amount of learning* for the untrained reference value condition for each participant in the same way as in the previous section in Experiment 1. The only difference is that when comparing the predicted learning in the untrained reference value condition to the absolute learning in the trained reference value condition, one needs to use between-subject comparison. To perform this test, first, we computed the predicted amount of learning in the group trained with the high-reference-values (con. 73% and ori. 25°) using Equation 2 by simply multiplying the absolute learning scores of the participants at the high-reference-values with the ratio of their initial thresholds at the two reference values ($\frac{IT_{Con30}}{IT_{Con73}}$ in Experiment 2, and $\frac{IT_{Ori0}}{IT_{Ori25}}$ in Experiment 3). We compared these predicted learning scores to the absolute learning scores of the observers in the low-reference-value conditions (con. 30% and ori. 0°) and found no difference between the two groups' scores (Experiment 2, contrast discrimination task: $t_{45} = 0.314$, $p = 0.755$, $d = 0.094$, Bayes Factor favoring no difference = 7.5; Experiment 3, orientation discrimination task: $t_{28} = 0.596$, $p = 0.556$, $d = 0.225$, Bayes Factor favoring no difference = 4.6, Figure 4, subpanel A in all panels).

Second, we computed the predicted amount of learning in the group trained with the **low**-reference-values (con. 30% and ori. 0°) derived from Equation 1 by solving it for $PL_{@HighRef}$,

$$PL_{@HighRef} = PL_{@LowRef} / \frac{IT_{@LowRef}}{IT_{@HighRef}}. \quad (3)$$

Using Equation 3, we divided the absolute learning scores of the participants at the low-reference-values with the ratio of their initial thresholds at the two reference values ($\frac{IT_{Con30}}{IT_{Con73}}$ in Experiment 2, and $\frac{IT_{Ori0}}{IT_{Ori25}}$ in Experiment 3). When comparing these predicted learning scores to the absolute learning scores of the observers in the high-reference-value conditions (con. 73% and ori. 25°), we found again no difference between the two groups' scores (Experiment 2, contrast discrimination task: $t_{45} = 0.689$, $p = 0.494$, $d = 0.206$, Bayes Factor favoring no difference = 5.7; Experiment 3, orientation discrimination task: $t_{28} = 1.091$, $p = 0.284$, $d = 0.412$, Bayes Factor favoring no difference = 2.9, Figure 4, subpanel B in all panels). In the Supplementary material, we provide further explanation as to why these between-subject comparison results support our claim that the amount of learning in these perceptual learning tasks was modulated only by the participants' perceptual scaling function without any additional processes.

Analyzing individual differences in initial thresholds and learning

We analyzed the individual differences within conditions and investigated how much of the inter-subject variability in learning could be explained by the initial discrimination threshold levels of the observers assuming a proportional relationship between initial performance and learning.

The individual differences in initial performance levels could explain a large part of the variability in learning in all experiments (variance explained in Experiment 1 was 20%, in Experiment 2 was 55%, and in Experiment 3 was 74%, Figure 3, subpanel D in all panels). To test whether the relationship between initial thresholds and the amount of learning was proportional, we computed the *relative learning* scores of the observers as the ratio of the initial and the posttraining discrimination thresholds ($\frac{\text{initial threshold}}{\text{final threshold}}$). If the relationship between the amount of learning and the initial discrimination threshold levels is strictly proportional, the relative learning scores should be the same at different initial threshold levels. Consequently, there should be no correlation between the relative learning scores and the initial threshold levels. Specifically, $PRE - POST(learning) = c \times PRE$ with a constant c.

Solving this equation for relative learning yields $\frac{PRE}{POST} = \frac{1}{1-c}$, which is a constant again. Following this analysis, in Experiment 1 we found that the positive correlation between learning and the initial thresholds completely disappeared when we used relative learning instead of the absolute learning scores. This suggests that the observers' learning was strictly proportional to their initial discrimination thresholds (Figure 4, top panel, C, and Table 1). In contrast to Experiment 1, in Experiment 2 and 3 a significant positive relationship between the relative learning and the initial thresholds remained suggesting that the amount of learning in these experiments was not strictly proportional to the initial threshold levels at the intersubject variability level (Figure 4, middle & bottom panels, subpanel C, and Table 1). On the one hand, this suggests that the relationship between learning and initial performance does not solely reflect the observers' perceptual scaling function from physical to perceptual magnitudes, but there are additional unknown factors strengthening that relationship beyond proportionality. Intersubject variability is known to reflect arousal level, attention, and motivation (Fahle & Henke-Fahle, 1996; Weiss, Edelman, & Fahle, 1993), each of which can influence the initial discrimination thresholds and can also be modulated by the training, causing a positive relationship between learning and initial performance. On the other hand, the correlations were much smaller between the initial discrimination thresholds and the relative learning than between the initial discrimination thresholds and the absolute learning. In Experiment 2,

Experiment	Correlation coefficient, r	95% CI	p Value
Correlation between initial thresholds and absolute learning (pre − post thresholds)			
Experiment 1 con-ws (Figure 3D, top panel)	0.45	0.12–0.69	0.007
Experiment 2 con-bs (Figure 3D, middle panel)	0.74	0.57–0.85	0.001
Experiment 3 ori-bs (Figure 3D, bottom panel)	0.86	0.72–0.93	0.001
Correlation between initial thresholds and relative learning (pre/post thresholds)			
Experiment 1 con-ws (Figure 4C, top panel)	0.05	−0.30–0.40	0.760
Experiment 2 con-bs (Figure 4C, middle panel)	0.31	0.01–0.55	0.035
Experiment 3 ori-bs (Figure 4C, bottom panel)	0.34	−0.03–0.63	0.061

Table 1. Analyzing intersubject variability with correlation. *Notes*: con-ws = contrast discrimination with within-subject design; con-bs = contrast discrimination with between-subject design; ori-bs = orientation discrimination with between-subject design.

the correlations were, $r = 0.74$ with absolute learning and $r = 0.31$ with relative learning, with a significant difference between them ($z = 2.911$, $p = 0.004$). In Experiment 3 the same correlations were $r = 0.86$ with absolute learning and $r = 0.34$ with relative learning with an even more significant difference between the two ($z = 3.625$, $p < 0.001$). This means that even when looking at intersubject variability, the relationship between learning and initial performance mainly reflects the effect of the perceptual scaling function of the observers. When the influence of the perceptual mapping is factored out by using the relative learning, most of the positive relationship disappears, and the explained variance drastically decreases (approximately from 70% to 10%, see the exact correlation coefficients above, and in Table 1).

To sum up our findings, the amount of learning was proportional to the initial threshold levels reflecting the effect of the observers' perceptual scaling function linking physical and perceptual magnitudes. This effect fully captured the observed relationship found in the within- and between-subject analyses when comparing the group means of the conditions with different stimulus base-intensities, and it also explained most of the individual variation between the participants within conditions.

Learning and generalization

The second goal of our study was to investigate whether the extent of generalization is proportional to the amount of learning in our paradigms. To this end, we analyzed intersubject variability and found positive correlations between the amount of learning and the extent of generalization in all of the experiments (see Supplementary Table S1, and Supplementary Figure S4).

Because the intersubject variability was much smaller when the reference orientation was at the cardinal orientation compared to the variability at 25°, the above correlational analysis could be misleading

due to the large differences in the variances of the learning and generalization scores (see Supplementary Figure S4 G and H). Therefore, we included two additional groups of observers in the orientation discrimination experiment. The observers underwent the same experimental protocol except that one group practiced with reference orientation 15° and the generalization of learning was assessed at 45°, while the other group practiced with 45° reference value and the transfer of learning was measured at 15° (see Supplementary materials and Supplementary Figure S3). In these groups, the intersubject variability was similar at both reference orientations (45° and 15°) and it was also large enough to study correlation between generalization and learning (Supplementary Figure S4 I and J).

Beside the positive relationship between learning and generalization, we also found positive correlations between the initial threshold levels and the amount of generalization in all experiments (see Supplementary Table S1, and Supplementary Figure S4). Since the measurement of generalization is also based on the estimation of the discrimination thresholds, observers' perceptual scaling function from physical to perceived magnitudes should influence the amount of generalization in the same way as it influences the amount of learning (see Figure 1 for explanation). This would automatically imply a positive relationship between initial threshold levels and the extent of generalization. However, we were interested in the relationship between learning and generalization without the obvious common influence of the initial discrimination thresholds. Therefore, we computed the partial correlations between learning and generalization while controlling for the initial threshold levels. Despite factoring out the effect of the initial thresholds, we found positive correlations in all experiments (Table 2, and see Methods, Analysis for more information about partial correlation). These findings validate our results, suggesting a positive relationship between the amount of learning and generalization in all experiments and confirms that the observed correlations were not due to

Experiment	Correlation coefficient, r	95% CI	p Value
Partial correlation between learning and absolute generalization while controlling for initial thresholds			
Experiment 1 transfer from con. 30% to con. 47%	0.65	0.23–0.87	0.004
Experiment 1 transfer from con. 73% to con. 47%	0.46	0.05–0.77	0.043
Experiment 2 transfer from con. 30% to con. 73%	0.46	0.00–0.76	0.050
Experiment 2 transfer from con. 73% to con. 30%	0.20	−0.30–0.61	0.366
Experiment 2 transfer from con. 30% to con. 47%	0.73	0.40–0.89	0.001
Experiment 2 transfer from con. 73% to con. 47%	0.55	0.11–0.81	0.015
Experiment 3 transfer from ori. 0° to ori. 25°	0.59	−0.03–0.88	0.020
Experiment 3 transfer from ori. 25° to ori. 0°	0.14	−0.52–0.62	0.622
Experiment 3 transfer from ori. 45° to ori. 15°	0.25	−0.43–0.75	0.545
Experiment 3 transfer from ori. 15° to ori. 45°	0.89	0.62–0.97	0.001
Partial correlation between learning and relative generalization while controlling for initial thresholds			
Experiment 1 transfer from con. 30% to con. 47%	−0.24	−0.56–0.29	0.356
Experiment 1 transfer from con. 73% to con. 47%	0.05	−0.45–0.53	0.848
Experiment 2 transfer from con. 30% to con. 73%	−0.14	−0.57–0.35	0.521
Experiment 2 transfer from con. 73% to con. 30%	−0.25	−0.64–0.25	0.259
Experiment 2 transfer from con. 30% to con. 47%	−0.20	−0.61–0.29	0.416
Experiment 2 transfer from con. 73% to con. 47%	−0.45	−0.76–0.02	0.053
Experiment 3 transfer from ori. 0° to ori. 25°	0.42	−0.26–0.82	0.116
Experiment 3 transfer from ori. 25° to ori. 0°	0.33	−0.36–0.78	0.233
Experiment 3 transfer from ori. 45° to ori. 15°	−0.05	−0.64–0.58	0.891
Experiment 3 transfer from ori. 15° to ori. 45°	−0.10	−0.54–0.67	0.761

Table 2. Top: Partial correlations between learning and absolute generalization. Bottom: Partial correlations between learning and relative generalization computed as generalization divided by the amount of learning. *Notes*: Transfer from con. 30% to con. 47% denotes the condition in which training were at reference value con. 30% and generalization was measured at con. 47%. The notations for the other conditions follow the same logic.

the self-evident modulating effect of the initial discrimination thresholds.

We also computed the relative generalization for each observer by taking the ratio of the extent of generalization and the amount of learning. If generalization is proportional to the amount of learning, this relative generalization should be constant at different amounts of learning because the proportionality relationship claims that *generalization* $= c \times learning$; thus, $\frac{generalization}{learning} = c$, where c is a constant. Indeed, using relative generalization the positive correlations, we found with the absolute generalization scores vanished and became statistically indistinguishable from zero (Table 2).

One potential caveat with this analysis is related to the fact that generalization was assessed by comparing the performance in the untrained conditions at pre- and posttest. If there were no learning between day two (first day of practice) and five (posttest) it would raise the possibility that the measured generalization scores mainly reflect the influence of the pretest which cannot be considered as true generalization because it is identical for the trained and untrained conditions. Indeed, looking at the learning curves in Figure 1 subpanels A, it is evident that most learning took place from Day 1 one to Day 2 in most experiments.

However, our analyses revealed that there was still a significant improvement in most of the conditions after the second day of practice (see Supplementary materials and Supplementary Figure S5). This means that the measurement of generalization used in the present study truly assesses generalization, even if it most probably overestimates somewhat its magnitude. Based on these measurements, our data support the claim that the extent of the generalization in our experiments was proportional to the observers' learning.

General discussion

In three experiments, we investigated (a) how initial performance, as quantified by discrimination threshold at pretest, and overall learning performance were related, and (b) how learning performance and ability to generalize were linked in customarily used perceptual learning tasks. Our goal was to identify general rules that apply to a wide range of conditions during perceptual learning. First, we confirmed the Weber-like law relationship between the initial threshold levels and the amount of learning reported by Astle et al. (2013) and showed that it essentially reflects the perceptual

scaling function of the observers without any evidence of additional learning-related processes. Moreover, we found that this proportionality relationship explained not only group mean results but also most of the individual variation across participants. Second, we found that the extent of generalization was proportional to the amount of observers' learning. In the following, we relate our results to the earlier literature and reflect on the implications of the present findings.

Initial performance and learning

First, we discuss the comparison of the low- and high-reference-value (i.e., stimulus base-intensity) conditions and how these results relate to the earlier findings of Astle et al. (2013). Second, we consider the results coming from the intersubject variability analysis and discuss its relation to previous studies (Aberg & Herzog, 2009; Astle et al., 2013; Fahle, 1997; Fahle & Henke-Fahle, 1996).

The results of Astle et al. (2013) and the current experiments are in agreement: They both show proportionally more learning in the conditions with higher initial thresholds compared to conditions with lower initial thresholds. Astle and his colleagues (2013) used a monocular, single-interval Vernier acuity task with a 10-day long training protocol, and they modulated the initial discrimination threshold levels by changing the eccentricity of the stimuli in a within-subject design. We applied binocular, two-interval contrast and orientation discrimination tasks with a five-day long training protocol and the initial discrimination threshold levels were modulated by changing the reference contrast and orientation values in within- and between-subject designs. Astle et al. (2013) also showed that the modulation of the initial performance level with crowding or with changing the size of the stimulus elicits the same effect on the amount of learning. The present study used different reference values to modulate initial performance, which again showed very similar effect on the amount of perceptual learning. Regardless of these differences, in both studies across six experiments, the amount of learning was proportional to the initial thresholds. Since these two studies found consistent results across three different paradigms under two different training protocols, by using different factors for modulating initial performance levels, together they point towards a general rule in perceptual learning that can predict the amount of learning from the initial discrimination threshold levels. Specifically, regardless of what mechanism constrains the visual discrimination thresholds the amount of learning will be proportional to the initial thresholds (Astle et al., 2013).

Regarding the origin of this proportionality rule, Astle and his colleagues' interpretation is quite different from ours. They proposed that the same cortical factors that put a limit on visual perception determining the discrimination thresholds constrain the amount of learning resulting in a Weber-like law during perceptual learning. We found that there was no extra constraint by any cortical factor that modulated learning in addition to the known perceptual processes. Rather, when perceptual scaling was considered at the individual level, the Weber-like law between initial thresholds and learning naturally emerged without any further assumptions. This result implies that, after the transformation of the input from the stimulus space to perceptual space takes place, the same amount of perceptual learning occurs at all stimulus intensity levels for all lower level visual attributes. Furthermore, the proportional relationship between initial thresholds and learning also implies that there was no change in the shape of the observer's perceptual scaling function due to the training; only the resolution got higher at the practiced stimulus intensities (i.e., the perceptual discrimination threshold decreased).

In principle, the proportional relationship between initial threshold and amount of learning could also be explained as a result of a particular combination of change in the shape of the perceptual scaling function and/or additional learning effects beyond the simple perceptual scaling that we suggest here. However, based on parsimony, we find such a complex explanation unlikely.

Considering intersubject variability, the amount of learning in our first experiment using a within-subject design was strictly proportional to the individual initial threshold levels in accordance with the results of Astle et al. (2013). However, in our other two experiments using a between-subject design, the amounts of learning increased more rapidly as a function of the initial threshold levels surpassing proportionality in line with the previous findings of Aberg and Herzog (2009) and Fahle and Henke-Fahle (1996). Exploring this discrepancy, we found that most of the variance in the relationship between initial discrimination threshold levels and learning was captured by the proportionality rule in all of our experiments. Therefore, while other (unknown) factors could also influence the relationship between initial threshold and learning, those represent only secondary effects. We attribute the origins of those secondary, unknown factors to arousal level, attention, and to motivation (Fahle & Henke-Fahle, 1996; Weiss et al., 1993), which can influence the initial discrimination thresholds and can also change due to practice, hence causing a deviation from the strict proportionality rule. Thus, intersubject variability can also be well explained by the proportional relationship between initial thresholds and learning.

Learning and generalization

Considering the link between the amount of learning and the extent of generalization, our results suggest that more learning predicts proportionally more generalization in the standard perceptual learning paradigms with five-day training. This proportionality relationship was supported by (a) the significant positive partial correlations between the amount of learning and the extent of *absolute generalization* while controlling for different initial threshold levels, and (b) by the nonsignificant partial correlations between the amount of learning and the extent of *relative generalization* (while controlling for initial threshold levels).

We can reconcile the contradiction between this conclusion and earlier reports showing more learning but less generalization after longer training (Hussain et al., 2012) by considering the two components of learning mentioned in the introduction: the specific characteristics of the training data and overfitting. Depending on whether or not the training data represents the space of the task well, acquiring more knowledge about this training set can help with generalization or hinder it. However, adjusting the internal model of the learner excessively to a training set regardless of how well it represents the space of the task (i.e., overfitting the data) will necessarily lead to less generalization. The interplay between these two components in the specific setup of Hussain et al. (2012) and Jeter et al. (2010) led to a lack of generalization. This effect might have been due to the increased training length applied in the tasks of Jeter and colleagues' study (2010) since encountering more training trials from the same kind increases the chances of overfitting (observers adjust their internal model more tightly according to the frequently observed trials). In contrast, the training length (in number of trials) in our experiments was fixed at about half of that used in the longest session of Jeter et al. (2010), implying less overfitting and more generalization. Therefore, our training protocol might have created a condition that did not overrepresent particular aspects of the space of the task as much as in previous studies (Hussain et al., 2012; Jeter et al., 2010), leading to the observation that the more observers learned, the more they generalized. Since a number of factors related to the task and the stimuli can influence when overfitting begins, the nature of specificity or transfer of learning might not be related to the amount of learning directly, but rather to the balance between the extent of learning, stimulus variability, and the given task with its specific features (Hussain et al., 2012; Jeter et al., 2010).

Clearly, this hypothesis of ours, suggesting that it is the stronger overfitting and not the larger amount of learning per se that is responsible for specificity in standard perceptual learning tasks, remains to be tested in future studies.

Conclusion

The present study investigated two simple, but general rules that can predict performance in perceptual learning paradigms. First, we confirmed that initial performance and learning are related through a Weber-like relationship regardless of the learning task and showed that this link is a direct consequence of the observers' perceptual scaling function relating physical intensities to perceived magnitudes. Second, we found that the more people learn under the typical five-day training protocol, the more they generalize. This implies that enhanced specificity reported in some previous studies were not an inherent consequence of the paradigm of perceptual learning with repetitive training but rather of overfitting the training set which is determined by a number of additional factors of the experimental design.

Acknowledgments

References

Aberg, K. C., & Herzog, M. H. (2009). Interleaving bisection stimuli—randomly or in sequence—does not disrupt perceptual learning, it just makes it more difficult. *Vision Research*, *49*(21), 2591–2598.

Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3):3, 1–11, https://doi.org/10.1167/12.3.3. [PubMed] [Article]

Adini, Y., Sagi, D., & Tsodyks, M. (2002, February

14). Context-enabled learning in the human visual system. *Nature*, *415*(6873), 790–793.

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *Journal of Vision*, *4*(12):2, 993–1005, https://doi.org/10.1167/4.12.2. [PubMed] [Article]

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

Ahissar, M., Merav, A., & Shaul, H. (1997, May 22). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Astle, A. T., Li, R. W., Webb, B. S., Levi, D. M., & McGraw, P. V. (2013). A Weber-like law for perceptual learning. *Scientific Reports*, *3*: 1158, https://doi.org/10.1038/srep01158.

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

Burton, G. J. (1981). Contrast discrimination by the human visual system. *Biological Cybernetics*, *40*(1), 27–38.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Lawrence Erlbaum Associates.

Cohen, Y., Daikhin, L., & Ahissar, M. (2013). Perceptual learning is specific to the trained structure of information. *Journal of Cognitive Neuroscience*, *25*(12), 2047–2060.

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.

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Research*, *37*(14), 1885–1895.

Fahle, M., & Henke-Fahle, S. (1996). Interobserver variance in perceptual performance and learning. *Investigative Ophthalmology & Visual Science*, *37*(5), 869–877.

Fechner, G. T. (1860). *The elements of psychophysics*. Leipzig, Germany: Breitkopf und Härtel (Reprinted, Bristol: Thoemmes Press, 1999).

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–1881.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, UK: John Wiley.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY:

Springer. https://doi.org/10.1007/978-0-387-84858-7.

Herzog, M. H., & Manfred, F. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141.

Hung, S.-C., & Seitz, A. R. (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(25), 8423–8431.

Hussain, Z., Bennett, P. J., & Sekuler, A. B. (2012). Versatile perceptual learning of textures after variable exposures. *Vision Research*, *61*, 89–94.

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

Jeter, P. E., Dosher, B. A., 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. [PubMed] [Article]

Karmali, F., Chaudhuri, S. E., Yi, Y., & Merfeld, D. M. (2016). Determining thresholds using adaptive procedures and psychometric fits: Evaluating efficiency using theory, simulations, and human experiments. *Experimental Brain Research*, *234*(3), 773–789.

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*(11), 4966–4970.

Kuai, S.-G., Shu-Guang, K., Jun-Yun, Z., Klein, S. A., Levi, D. M., & Cong, Y. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience*, *8*(11), 1497–1499.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research*, *21*(4), 457–467.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2; Suppl. 2), 467.

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10):13, 1–14, https://doi.org/10.1167/12.10.13. [PubMed] [Article]

Mansfield, R. J. W. (1974, December 20). Neural basis of orientation perception in primate vision. *Science*, *186*(4169), 1133–1135.

Mikellidou, K., Cicchini, G. M., Thompson, P. G., &

Burr, D. C. (2015). The oblique effect is both allocentric and egocentric. *Journal of Vision*, *15*(8): 24, 1–10, https://doi.org/10.1167/15.8.24. [PubMed] [Article]

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Research*, *24*(2), 121–128.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

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.

Polat, U., Schor, C., Tong, J.-L., Zomet, A., Lev, M., Yehezkel, O., . . . Levi, D. M. (2012). Training the brain to overcome the effect of aging on the human eye. *Scientific Reports*, *2*: 278.

Regan, D., & Price, P. (1986). Periodicity in orientation discrimination and the unconfounding of visual information. *Vision Research*, *26*(8), 1299–1302.

Ross, H. E., & Murray, D. J. (1996). *E. H. Weber on the tactile sense*. Hove, UK: Erlbaum, Taylor & Francis.

Rouder, J. N., Speckman, P. L., Dongchu, S., Morey, R. D., & Geoffrey, I. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

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*(Pt. 3), 797–810.

Seitz, A. R., & Watanabe, T. (2003, March 6).

Psychophysics: Is subliminal learning really passive? *Nature*, *422*(6927), 36.

Solomon, J. A., & Tyler, C. W. (2017). Improvement of contrast sensitivity with practice is not compatible with a sensory threshold account. *Journal of the Optical Society of America*, *34*(6), 870–880.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.

Streiner, D. L. (2003). Unicorns do exist: A tutorial on "proving" the null hypothesis. *Canadian Journal of Psychiatry*, *48*(11), 756–761.

Wang, R., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2014). Vernier perceptual learning transfers to completely untrained retinal locations after double training: A "piggybacking" effect. *Journal of Vision*, *14*(13):12, 1–10, https://doi.org/10.1167/14.13.12. [PubMed] [Article]

Weiss, Y., Edelman, S., & Fahle, M. (1993). Models of perceptual learning in Vernier hyperacuity. *Neural Computation*, *5*(5), 695–718.

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: CB*, *18*(24), 1922–1926.

Yehezkel, O., Sterkin, A., Lev, M., Levi, D. M., & Polat, U. (2016). Gains following perceptual learning are closely linked to the initial visual acuity. *Scientific Reports*, *6*: 25188.

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

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. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(37), 12323–12328.