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

Gábor Lengyel* and József Fiser*
Central European University

* Department of Cognitive Science, Central European University, Október 6 u. 7, Budapest 1051, Hungary

Correspondence: Gábor Lengyel (lengyel.gaabor@gmail.com), József Fiser (fiserj@ceu.edu)

Supplementary material

Quantifying the decrease in thresholds and lapse rates after practice

Estimating discrimination thresholds using adaptive staircase methods confounds errors due to lapses (lack of attention) with errors due to real perceptual indiscriminabilty (Solomon & Tyler, 2017). Although the 3-down-1-up staircase can be robust to the initial attentional lapses (Karmali, Chaudhuri, Yi, & Merfeld, 2016) lapses are not necessarily limited to the initial trials in novice observers. We investigated this potential confound by first estimating the lapse rates and the discrimination thresholds for each observer in the pre- and posttests by fitting psychometric curves to the observers' performance and then, testing whether the thresholds and/or the lapse rates decreased due to learning. We confirmed that both participants' thresholds and attentional lapses decreased due to practice.

We fitted cumulative Weibull distributions (psychometric curves) to participants' data at the pre- and the posttests:

$$P(x) = \varepsilon + (1 - \varepsilon - \gamma)\,(1 - e^{-(x/\alpha)^{\beta}}) \qquad \text{(Equation S1)}$$

In this formula, x is the stimulus strength which is the contrast and the orientation difference in % contrast and in degrees respectively. $P(x)$ is the fraction of the correct responses at stimulus strength x. The parameters denote the following: α is the threshold, β is the slope, γ is the lapse rate, and ε is the chance performance level which was 0.5 in our tasks. The lapse rates and the thresholds of the observers were estimated by the best-fitting value of the γ and α parameters using maximum likelihood estimation.

The lapse rates decreased significantly due to training in the contrast discrimination with within-subject design (t_{16}= 3.154, p=0.006, d=0.786, Fig. S1, top row, in the middle) and in the orientation discrimination experiments (t_{29}= 3.226, p=0.003, d=0.599, Fig. S1, top row, on the right), however it did not change in the contrast experiment with between-subject design (t_{30}= 0.282, p=0.779, d=0.052, Fig. S1, top row, on the left).

The thresholds of the participants decreased significantly in all experimental condition after training. In the orientation discrimination experiment we obtained t_{14}= 5.834, p<0.001, d=1.559 with reference orientation 0°, and t_{14}= 3.319, p=0.005, d=0.882 with the reference orientation 25° (Fig. S1, middle row, on the right). In the contrast discrimination experiment with between-subject design we obtained t_{14}= 2.969, p=0.010, d=0.793, at reference contrast 30%, and t_{15}= 3.298, p=0.005, d=0.851 at reference contrast 73% (Fig. S1, middle row, on the left). In the contrast discrimination experiment with within-subject design we obtained t_{15}= 3.137, p=0.007, d=0.809, at reference contrast 30%, and t_{15}= 3.748, p=0.002, d=0.968 at reference contrast 73% (Fig. S1, middle row, in the middle).

We compared the two measurements assessing the decrease in the observers' thresholds due training. In all experiments there were large positive correlations between the decrease in thresholds estimated by the staircase and by the best-fitted Weibull function. These correlations were r= 0. 66, p<0.001, CI_{95}=0.40-0.83 in the contrast experiment with between-subject design (Fig. S1, bottom row, on the left), r= 0. 69, p<0.001, CI_{95}=0.45-0.84 in the contrast experiment with within-subject design (Fig. S1, bottom row, in the middle), and r= 0. 79, p<0.001, CI_{95}=0.59-0.90 in the orientation experiment with between-subject design (Fig. S1, bottom row, on the right).

Although the observers' thresholds decreased significantly due to learning even when we estimated with psychometric curves instead of the reversal points of the staircase method, this will not solve our problem of knowing whether the improvement reflects a decrease in thresholds or a decrease in lapse rates. A reduction in the lapse rate would shift the whole psychometric curve up which in many cases would also result in a decrease in the value of the threshold parameter. To test whether there was any decrease in the thresholds beyond the decrease of the lapse rates, we computed a hypothetical psychometric curve for each subject by adding the amount of decrease in lapse rate due to training to each of the data point of the psychometric curve fitted to the pre-training performance. This method shifted the participants' pre-training psychometric curves up by as much as their lapse-rates decreased after the training and thus, this hypothetical psychometric curve represents approximately the improvement that would have been caused by only improving in lapse rates. We compared the thresholds of the post-training (best-fitting) true psychometric curves to the thresholds of the hypothetical (best-fitting) psychometric curves that assumes only lapse rate improvement. We found that thresholds after the training were significantly lower than the corresponding thresholds of the hypothetical psychometric curves that represented the threshold values had they been solely under the control of the decreases in the lapse rates (Fig. S2). In the orientation discrimination experiment we obtained t_{14}= 5.834, p<0.001, d=1.559 with reference orientation 0°, and t_{14}= 3.319, p=0.005, d=0.882 with the reference orientation 25° (Fig. S2, first row, on the right). In the contrast discrimination experiment with between-subject design we obtained t_{14}= 3.271, p=0.006, d=0.874, at reference contrast 30%, and t_{15}= 3.567, p=0.003, d=0.921 at reference contrast 73% (Fig. S2, first row, on the left). In the contrast discrimination experiment with within-subject design we obtained t_{15}= 2.709, p=0.016, d=0.699, at reference contrast 30%, and t_{15}= 3.857, p=0.002, d=0.996 at reference contrast 73% (Fig. S2, first row, in the middle). Furthermore, this threshold improvement that was controlled for the decrease in lapse rate significantly correlated with the threshold improvement measured by the reversal points from the staircase procedure. These correlations were r= 0. 58, p<0.001, CI_{95}=0.27-0.78 in the contrast experiment with between-subject design (Fig. S2, second row, on the left), r= 0. 68, p<0.001,

CI_{95}=0.43-0.84 in the contrast experiment with within-subject design (Fig. S2, second row, in the middle), and r= 0. 74, $p<0.001$, CI_{95}=0.50-0.87 in the orientation experiment with between-subject design (Fig. S2, second row, on the right).

These results suggest that the decrease in the thresholds after practice was not solely due to the decrease in the lapse rates, and this improvement in perception can be approximated by computing the geometric mean of the reversal points of the adaptive staircase procedure.

Extended explanation for the statistical method used in the result section

The ratio of the observer's initial discrimination thresholds used for scaling the learning scores ($\frac{IT_{Con30}}{IT_{Con73}}$, and $\frac{IT_{Ori0}}{IT_{Ori25}}$) characterizes the observer's individual perceptual scaling function at the two measured stimulus base-intensities. Therefore, in the first case (*Eq. 2*) the multiplication of the high-reference-value learning scores with participants' initial threshold ratios scaled **down** participants' learning with the extend of how much larger their initial discrimination thresholds were at the high reference values compared to the low reference values prior to the practice (Fig. 4, subpanel A in all panels). This quantity gave us the *predicted amount of learning* in the untrained low-reference-value condition which can be compared to the measured *absolute learning* in the other group practicing with that low-reference-value. If the proportionality rule captured by *Eq. 1* holds, the *predicted* low-reference-value learning scores should be indistinguishable from the *absolute* low-reference-value learning scores. Alternatively, if some additional processes influence learning beyond the observers' perceptual scaling, and the amount of learning will deviate from proportionality rule, the *predicted* low-reference-value learning scores should be significantly different from the *absolute* low-reference-value learning scores. In the second case (*Eq. 3*) the division of the low-reference-value learning scores with the participants' initial threshold ratios scaled **up** participants' learning score with the extend of how much smaller their initial discrimination thresholds were at the low stimulus intensity compared to those at the high intensity (Fig. 4, subpanel B in middle and bottom panels). This quantity gave us the *predicted amount of learning* in the untrained high-reference-value condition which can be compared to the measured *absolute learning* in the other group practicing with that high-reference-value. The logic of the comparison of the *predicted* high-reference-value learning scores to the *absolute* high-reference-value learning scores is the same as in the previous paragraph.

Orientation discrimination experiments

In the orientation discrimination experiments separate groups of observers were trained to discriminate around four different reference values: 0°, 15°, 25°, and 45°. Regarding the investigation of the relationship between initial performance and learning we used 15° and 45° reference values in the first orientation discrimination experiment which did not elicit significant difference in the initial discrimination thresholds. Consequently, we could not test the effect of the different initial performance levels on the amount of learning. In the second experiment we used 0° and 25° for the orientation references, and we found a large difference between the initial discrimination thresholds, which enabled us to investigate how initial thresholds modulates the amount of learning. In the main text we only reported the results of the latter orientation discrimination experiment. However, in order to show all of our data, we present here all the analysis that we used in the main text for the first orientation discrimination experiment too (in which observers practiced with either 15° or 45° reference values).

Although the initial thresholds seem higher at 45° than at 15° the difference did not reach significance (t_{20}= 1.500, p=0.149, d=0.670, Fig. S3A). There was significant perceptual learning in both conditions (p<0.05, Fig. S3B & C) but, the amount of learning did not differ in the two groups (t_{20}= 1. 499, p=0.150, d=0.670, Fig. S3B). We computed the *predicted* learning scores in the group which practiced with 45° reference value using *Eq. 2* as: $\mathrm{Learning}_{\mathrm{ori45}}(\mathrm{PRE\text{-}POST}) * \frac{\mathrm{PRE}_{\mathrm{ori15}}}{\mathrm{PRE}_{\mathrm{ori45}}}$ (see main *Results* for more information). When we compared the *predicted* learning to the absolute learning scores in the group which practiced with 15° reference orientation the difference was not significant (t_{20}= 1. 067, p=0.299, d=0.477, Fig. S3F). Similarly, the *predicted* learning in the group which practiced with the 15° reference values was computed using *Eq. 3* as: $\mathrm{Learning}_{\mathrm{ori15}}(\mathrm{PRE\text{-}POST})/\frac{\mathrm{PRE}_{\mathrm{ori15}}}{\mathrm{PRE}_{\mathrm{ori45}}})$ and it did not differ significantly from the absolute learning scores in the 45° reference group (t_{20}= 1. 217, p=0.238, d=0.544, Fig. S3G). In terms of the inter-subject variability, there was a large positive correlation between the amount of learning and the initial threshold levels (r= 0. 85, p<0.001, CI=0.66-0.94, Fig. S3D). The correlation between relative learning (PRE/POST thresholds) and the initial threshold levels was also significant (r= 0. 44, p=0.039, CI_{95}=0.12-0.73, Fig. S3E) but smaller than the correlation between absolute learning and the initial thresholds (z = 2.684, p = 0.007). These results are in line with the results and the conclusion of the main text.

Regarding the generalization of learning, the inter-subject variability was much smaller with reference orientation 0° than with all other reference orientations (see Fig. 3, bottom panels, dots in purple, and Fig. S4G & H). Therefore, in the second orientation discrimination experiment the correlations between the amount of learning and the extent of generalization gave an unreliable estimate of the true linear relationship between generalization and learning due to the large differences in the variances of the two random variables. Here the two random variables were (1) learning at 0° and generalization at 25°, and (2) learning at 25° and generalization at 0°. Thus, we used the first orientation discrimination experiment with 15° and 45° reference values in the analysis investigating the relationship between learning and generalization (see R*esults* in the main text for more information).

Analyzing the amount of learning from the second day on

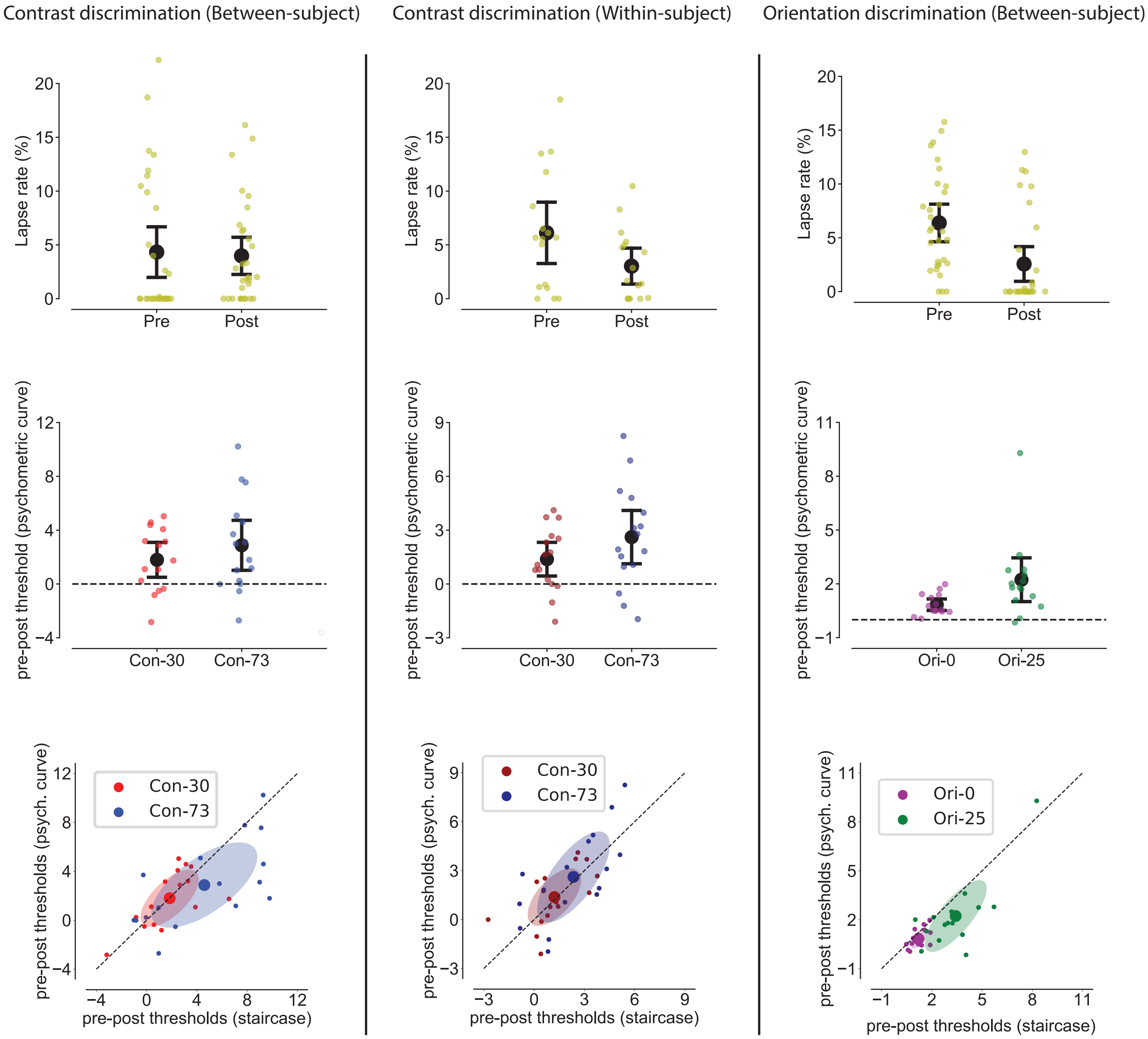
A potential problem weakening the measurement of generalization emerges when no learning took place from Day 2 to Day 5. In this case, learning in the untrained conditions (i.e. improvement at untrained reference values) does not necessary indicate generalization since the improvement in the trained conditions could be due to the pretest during which observers completed the same amount of trials in the trained and in the untrained conditions. To eliminate this problem, we tested whether there was further improvement in the experiments after the second day of practice and we found that there was significant learning after the second day in most of the conditions. Specifically, we found significant learning from Day 2 to Day 5 in the orientation discrimination experiments (Fig. S5, bottom panels) at reference orientation 15° (t_{10}==3.05, p=0.01), 45° (t_{10}==3.87, p=0.003), 25° (t_{14}==2.64, p=0.02), and non-significant learning at 0° (t_{14}==0.72, p=0.48). We also found significant, and marginally significant learning from Day 2 to Day 5 in most of the conditions of the contrast discrimination experiments (Fig. S5, top panels). Specifically, we obtained t_{23}==1.74, p=0.09 and t_{22}==2.57, p=0.01 in the contrast experiment with between-subject design at reference con. 30% and 73% respectively. In the contrast experiment with within-subject design we found t_{16}==0.37, p=0.71 and t_{16}==1.77, p=0.09 at reference con. 30% and 73% respectively.

There was no improvement after the second day in the cardinal (0°) reference orientation condition in the orientation discrimination task (Fig. S5, bottom right). However, we did not use this group of observers in the analysis for generalization of learning anyway because of its excessively small inter-subject variability (see *results, learning and generalization* and *supplementary materials, orientation discrimination experiments* for more detail). Regarding the contrast experiment with within-subject conditions (in which observers show the lowest amount of improvement from Day 2 across experiments) we cannot conclude that there was no further improvement from the Day 2 in this condition either. This is because there was a marginally significant improvement in the condition with reference contrast 73%, and only 3 subjects showed no improvement from Day 2 (Fig. S5, top left). Thus, learning could have transferred from that condition to the untrained middle reference value con. 47% in most participants.

References

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. Experimentelle Hirnforschung. Experimentation Cerebrale*, *234*(3), 773–789.

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.



Contrast discrimination (Between-subject)

Contrast discrimination (Within-subject)

Orientation discrimination (Between-subject)

Figure S1. Top row: The distribution of lapse rates at pre- (before training) and posttests (after training). **Middle row:** The difference between observers pre- and post-thresholds estimated by the threshold parameter (α) of the best-fitting psychometric curve. **Bottom row:** The improvement in discrimination thresholds due to learning using the reversal points from the adaptive staircase (x axis) is compared to the improvement in discrimination thresholds estimated by the threshold parameter (α) of the best-fitting psychometric curves of the participants at pre- and posttest (y axis). **Left column:** contrast discrimination task, between-subject design. **Middle column:** contrast discrimination task, within-subject design. **Right column:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean. Error ellipses show one standard deviation and the dashed lines mark the x=y values.

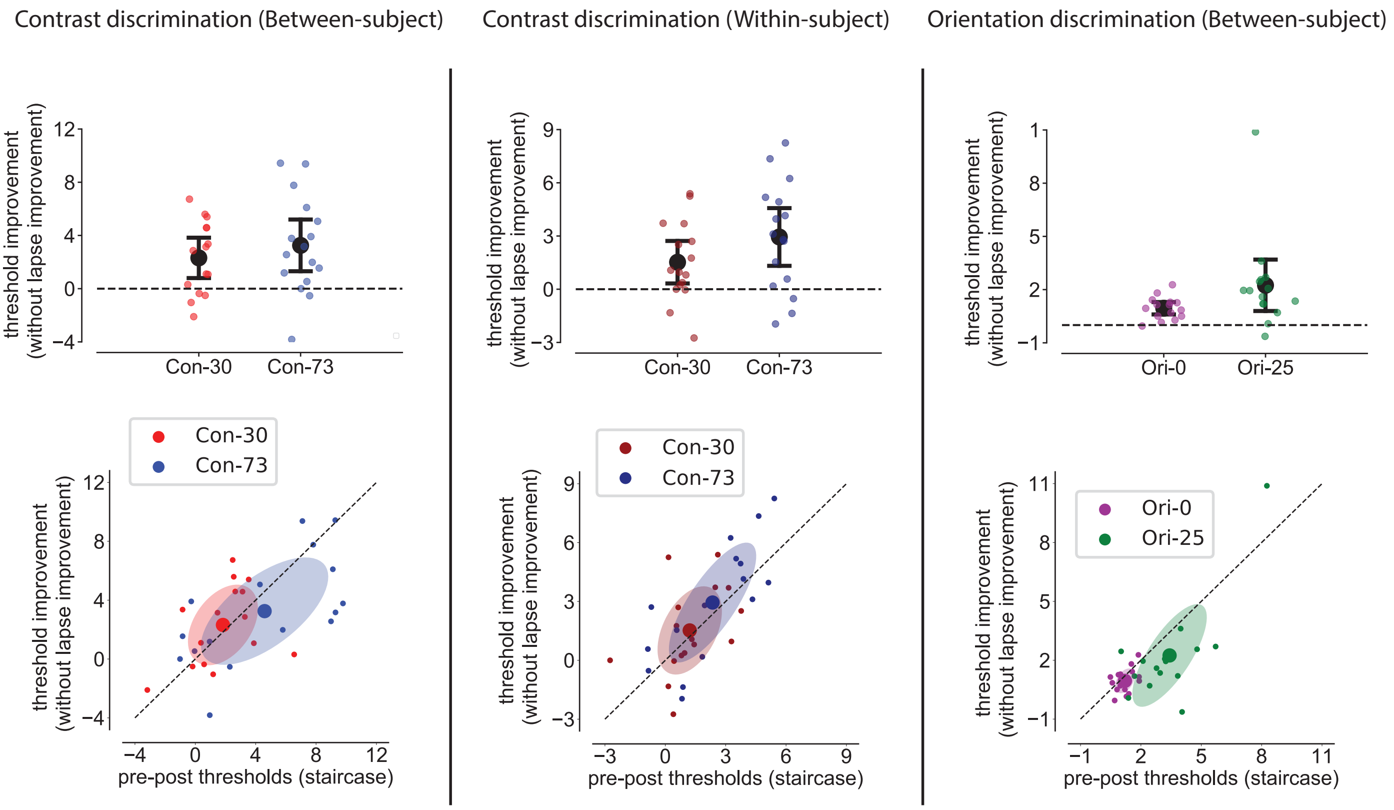


Figure S2. First row: The decrease in thresholds beyond the decrease in the lapse rates after the training (estimated by best-fitting psychometric curves). **Second row:** The improvement in discrimination thresholds due to learning using the reversal points from the adaptive staircase (x axis) is compared to the improvement in discrimination thresholds beyond the decrease of the lapse rates estimated by best-fitting psychometric curves of the participants using hypothetical performance assuming only lapse rate decrease due to training and participants post-training performance (y axis). **Left column:** contrast discrimination task, between-subject design. **Middle column:** contrast discrimination task, within-subject design. **Right column:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean. Error ellipses show one standard deviation and the dashed lines mark the x=y values.

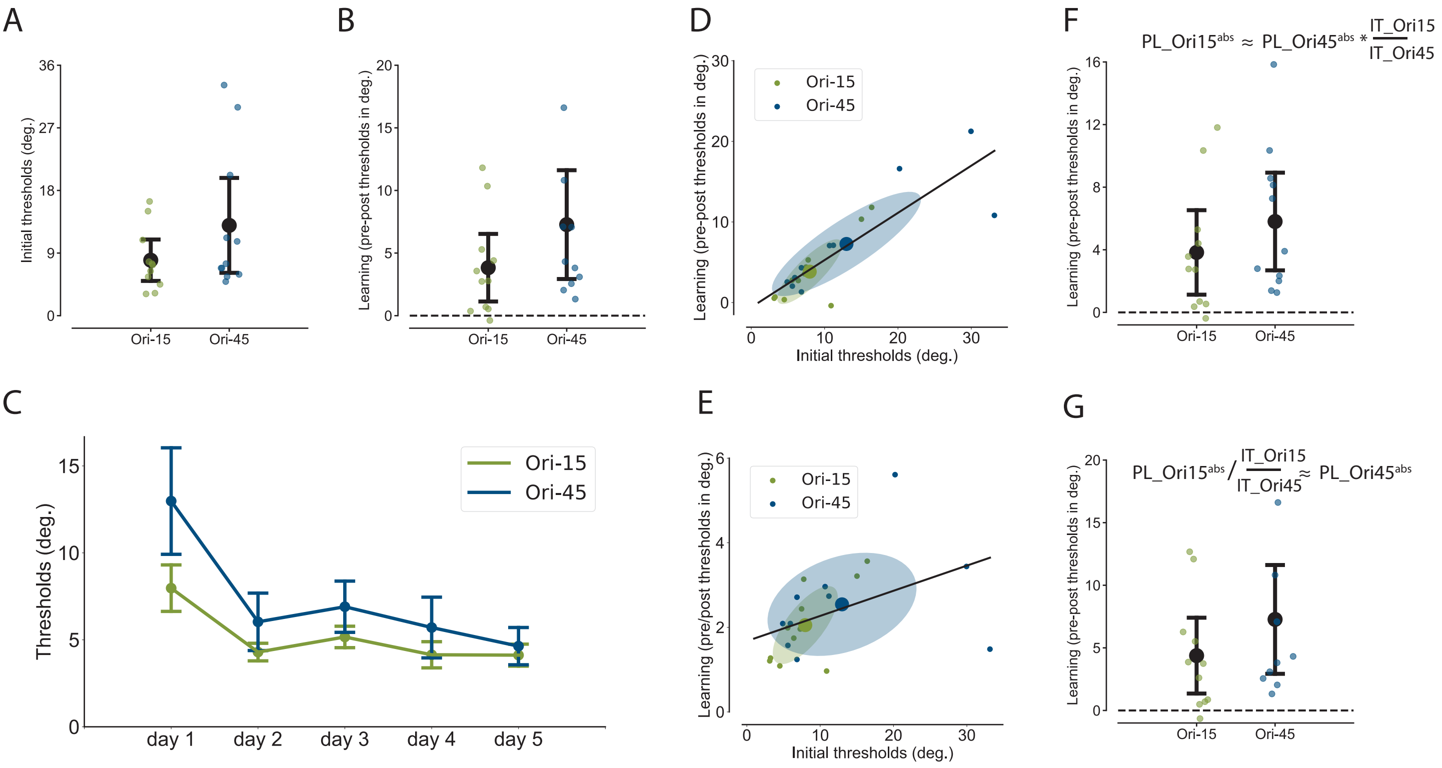


Figure S3. **(A)** Initial discrimination thresholds and **(B)** the amount of learning at the two measured reference values. **(A, B, F** & **G)** Error bars represent 95% confidence intervals on the mean. **(C)** Learning curves for the 5-day training protocol for the two measured reference values. Error bars show one SEM. **(D)** Learning as a function of initial discrimination thresholds. **(E)** Relative learning measured as initial discrimination thresholds divided by the post training thresholds as a function of the initial threshold levels. **(D** & **E)** Error ellipses show one standard deviation, and black lines show linear regression lines fitted to the points from both conditions. **(F)** Comparing the absolute learning in the low reference value condition (gold points) to the predicted learning in the high reference value condition (blue points). **(G)** Comparing the predicted learning in the low reference value condition (gold points) to the absolute learning in the larger reference value condition (blue points). **(F** & **G)** The equations above the error bars represent the functions of the scaling. *Learning15*, *-45* denotes the normal learning scores in the specified reference value condition. *init˙thr15*, *-45* denotes the initial thresholds at the specified reference values.

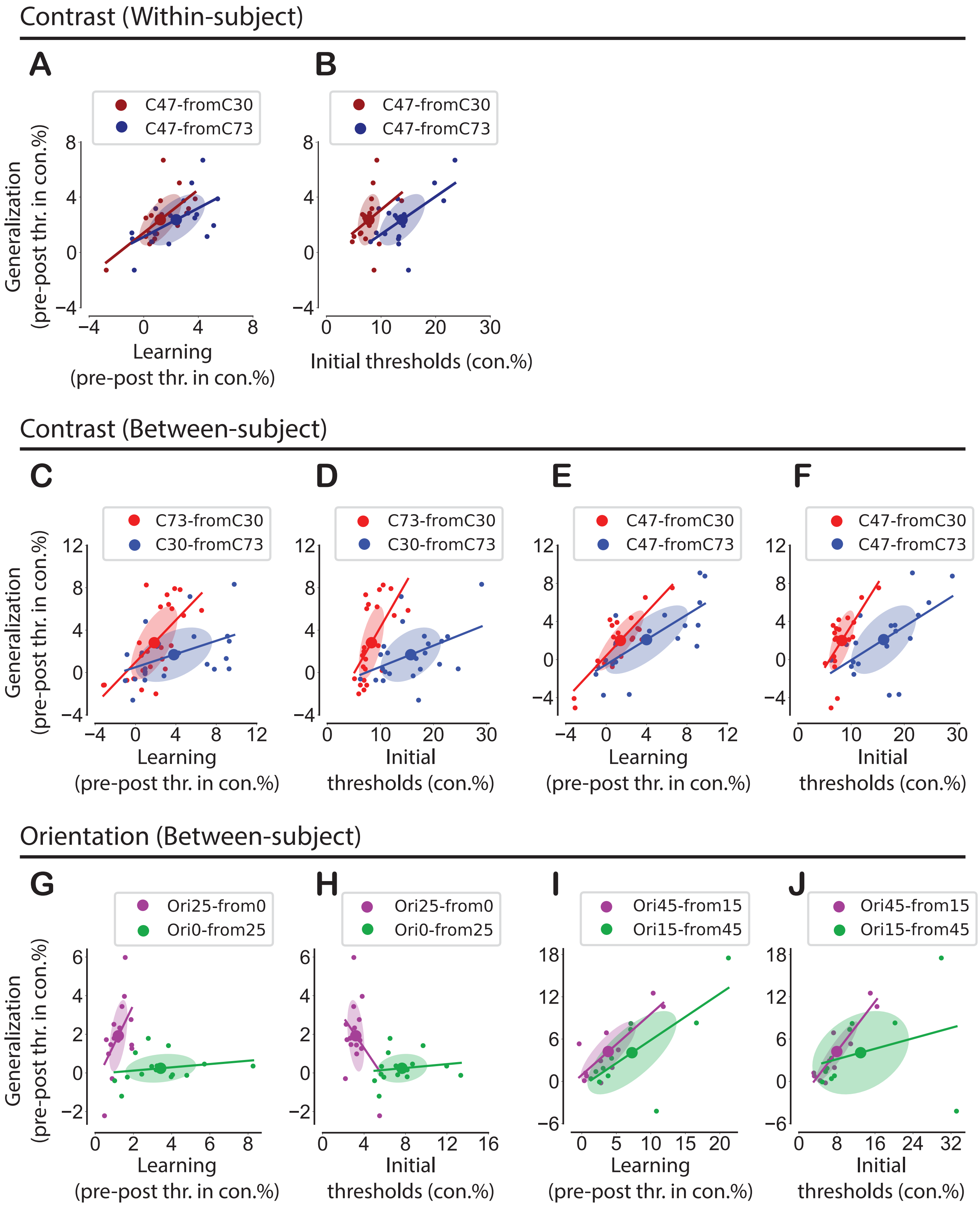


Figure S4. Top panel: contrast discrimination task, within-subject design. **Middle panel:** contrast discrimination task, between-subject design. **Bottom panel:** orientation discrimination task, between-subject design. **(A, C, E, G & I):** Generalization as a function of learning. **(B, D, F, H & J):** Generalization as a function of initial discrimination thresholds. In all plots error ellipses show one standard deviation and colored lines represent linear regression lines for the corresponding conditions. The first part of the labels (C73-, C30-, C47-, Ori0-, Ori25-) denotes the reference value at which the generalization was measured, while the second part of the labels (-fromC73, -fromC30, -from25, -from0) denotes the practiced reference values from which the learning transferred.

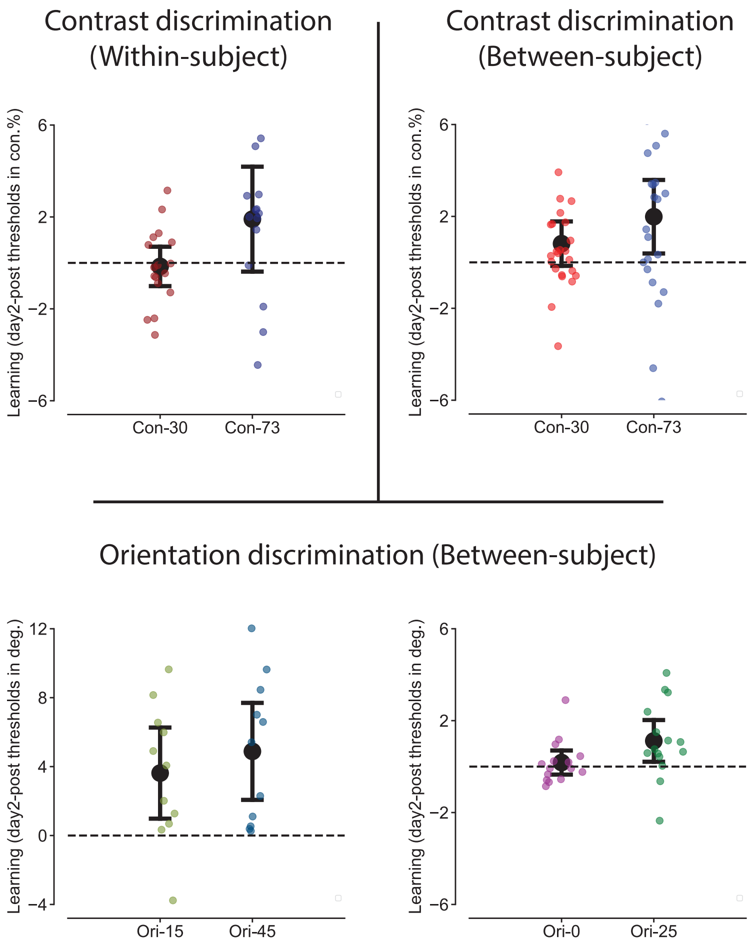


Figure S5. The amount of learning between Day 2 and the final posttest. **Top panel:** contrast discrimination task, within-subject design. **Middle panel:** contrast discrimination task, between-subject design. **Bottom panel:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean.

Correlations between learning and generalization			
Experiment	Correlation coefficient	95% Confidence interval	p value
Exp 1. transfer to con. 47% from con. 30% (Fig. S4 A, red line)	r = 0.64	CI95 = 0.23 - 0.87	p = 0.005
Exp 1. transfer to con. 47% from con. 73% (Fig. S4 A, blue line)	r = 0.58	CI95 = 0.13 - 0.84	p = 0.013
Exp 2. transfer to con. 73% from con. 30% (Fig. S4 C, red line)	r = 0.66	CI95 = 0.34 - 0.85	p < 0.001
Exp 2. transfer to con. 30% from con. 73% (Fig. S4 C, blue line)	r = 0.46	CI95 = 0.05 - 0.74	p = 0.027
Exp 2. transfer to con. 47% from con. 30% (Fig. S4 E, red line)	r = 0.85	CI95 = 0.63 - 0.94	p < 0.001
Exp 2. transfer to con. 47% from con. 73% (Fig. S4 E, blue line)	r = 0.74	CI95 = 0.43 - 0.90	p < 0.001
Exp 3. transfer to ori. 25° from ori.0° (Fig. S4 G, purple line)	r = 0.53	CI95 = 0.00 - 0.82	p = 0.044
Exp 3. transfer to ori.0° from ori.25° (Fig. S4 G, green line)	r = 0.21	CI95 = -0.35 - 0.66	p = 0.450
Exp 3. transfer to ori. 15° from ori.45° (Fig. S4 I, green line)	r = 0.82	CI95 = 0.40 - 0.95	p = 0.002
Exp 3. transfer to ori. 45° from ori.15° (Fig. S4 I, purple line)	r = 0.72	CI95 = 0.19 - 0.92	p = 0.013
Correlations between initial thresholds and generalization			
Exp 1. transfer to con. 47% from con. 30% (Fig. S4 B, red line)	r = 0.36	CI95 = -0.16 - 0.73	p = 0.154
Exp 1. transfer to con. 47% from con. 73% (Fig. S4 B, blue line)	r = 0.62	CI95 = 0.18 - 0.85	p = 0.008
Exp 2. transfer to con. 73% from con. 30% (Fig. S4 D, red line)	r = 0.61	CI95 = 0.26 - 0.82	p = 0.002
Exp 2. transfer to con. 30% from con. 73% (Fig. S4 D, blue line)	r = 0.44	CI95 = 0.02 - 0.73	p = 0.034
Exp 2. transfer to con. 47% from con. 30% (Fig. S4 F, red line)	r = 0.67	CI95 = 0.30 - 0.87	p = 0.002
Exp 2. transfer to con. 47% from con. 73% (Fig. S4 F, blue line)	r = 0.60	CI95 = 0.18 - 0.82	p = 0.008
Exp 3. transfer to ori. 25° from ori.0° (Fig. S4 H, purple line)	r = -0.35	CI95 = -0.74 - 0.21	p = 0.20
Exp 3. transfer to ori.0° from ori.25° (Fig. S4 H, green line)	r = 0.16	CI95 = -0.40 - 0.63	p = 0.565
Exp 3. transfer to ori. 15° from ori.45° (Fig. S4 J, green line)	r = 0.91	CI95 = 0.68 - 0.98	p = 0.002
Exp 3. transfer to ori. 45° from ori.15° (Fig. S4 J, purple line)	r = 0.32	CI95 = -0.37 - 0.78	p = 0.341

Table S1. Analyzing the linear relationship between the extent of generalization, the amount of learning, and the initial discrimination thresholds.