1 **Variability in training unlocks generalization in visual perceptual learning through invariant**

2 **representations**

3 Giorgio L. Manenti[1,2,3], Aslan Satary Dizaji[1], Caspar M. Schwiedrzik[1,2*†]

[1]Neural Circuits and Cognition Lab, European Neuroscience Institute Göttingen – A Joint Initiative of the University Medical Center Göttingen and the Max Planck Society, Grisebachstraße 5, 37077 Göttingen, Germany; [2]Perception and Plasticity Group, German Primate Center – Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany; [3]Systems Neuroscience Program; Graduate School for Neurosciences, Biophysics and Molecular Biosciences (GGNB), Göttingen, Germany

4

5 *Correspondence: c.schwiedrzik@eni-g.de

6 † Lead contact

7   **Summary**

8   Stimulus and location specificity are long considered hallmarks of visual perceptual learning. This

9   renders visual perceptual learning distinct from other forms of learning, where generalization can be

10   more easily attained, and unsuitable for practical applications, where generalization is key. Based on

11   hypotheses derived from the structure of the visual system, we test here whether stimulus variability

12   can unlock generalization in perceptual learning. We train subjects in orientation discrimination, while

13   we vary the amount of variability in a task-irrelevant feature, spatial frequency. We find that

14   independently of task difficulty, this manipulation enables generalization of learning to new stimuli

15   and locations, while not negatively affecting the overall amount of learning on the task. We then use

16   deep neural networks to investigate how variability unlocks generalization. We find that networks

17   develop invariance to the task-irrelevant feature when trained with variable inputs. The degree of

18   learned invariance strongly predicts generalization. A reliance on invariant representations can explain

19   variability-induced generalization in visual perceptual learning, suggests new targets for

20   understanding the neural basis of perceptual learning in high-order visual cortex, and presents an easy

21   to implement modification of common training paradigms that may benefit practical applications.

22

23

24   **Keywords:** visual perceptual learning, orientation discrimination, generalization, deep neural

25   networks

2

26 **Introduction**

27 A fundamental problem for perception is to extract reliable information from a highly variable signal[1].

28 It is widely accepted that the visual system achieves this by learning what is consistent in its inputs, a

29 process called *perceptual learning* (PL). However, how the enormous variability in the environment[2]

30 impacts PL *itself* is not well understood. Here, we ask how the visual system solves the challenge of

31 variability for learning.

32 Variability poses both a problem and an opportunity for PL. Varying stimuli are problematic because

33 they entail reduced predictability[3], weaker memory traces[4], uncertain reward assignment[5], and non-

34 linear decision rules[6]. They can thus slow down or impair learning. Consequently, state-of-the-art

35 research and commonly employed PL protocols aim to maximize PL using highly unnatural conditions

36 where variability is minimized, e.g., when learning one stimulus alternative at a time. However, what

37 is often overlooked is that variability can be a great asset for learning as it facilitates *generalization*

38 (i.e., applying learned behavior to new stimuli): variability may foster the extraction of core features

39 across stimuli through abstraction, concept learning, and rule derivation. In many learning domains,

40 variability acts as a catalyst for generalization[7]. In fact, theoretical studies suggest that the degree of

41 variability during learning determines whether the system specializes on specific stimuli by

42 memorizing them or instead learns generalizable rules[8]. Although generalization is frequently

43 considered the ultimate goal of learning[9], the current mainstream view on PL disregards variability

44 and generalization. Instead, PL is often said not to generalize[10,11]. This renders visual PL an outlier

45 amongst many learning phenomena.

46
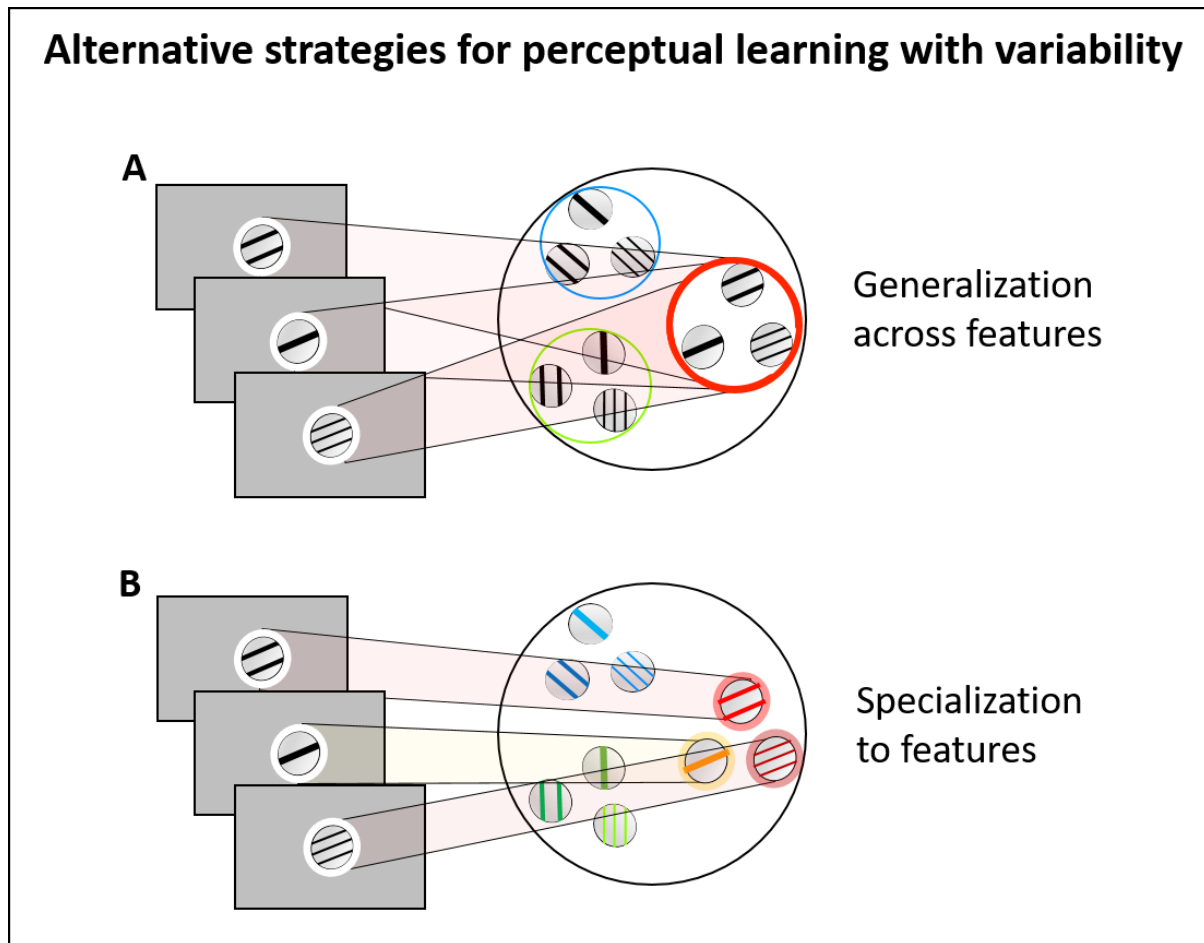
Figure 1: Alternative strategies for visual perceptual learning with variability. A. Generalization strategy: if input varies from trial to trial, learning could rely on neurons invariant to task-irrelevant features of the stimuli. These neurons deal with variability by subsuming it in their broad tuning and can accommodate generalization to new stimuli. However, their tuning to the task-relevant feature may not be very precise. **B**. Specialization strategy: alternatively, the system could implement learning with neurons narrowly tuned to task-relevant and task-irrelevant features. These neurons can provide high precision information for the task at hand. However, more neurons need to undergo plasticity than in A if inputs vary. Moreover, a specialization strategy does not lead to generalization, as new, untrained neurons are needed to accommodate new inputs.

Yet, the structure of the visual system suggests (at least) two ways to solve the challenge of variability during PL: the system could *generalize* by relying on invariant representations or *specialize* on both task-relevant and -irrelevant aspects of the stimuli, using highly precise neurons narrowly tuned to both dimensions. If the visual system employs any of these strategies is currently unknown.

A generalization strategy based on invariant representations, e.g., in ventral temporal cortex (VTC), deals with variability by subsuming it: invariant representations provide information about visual

64   stimuli in an abstract form, irrespective of task-irrelevant variability in low-level details of the input

65   (Fig. 1A). Thus, even if stimuli differ in appearance from trial to trial, invariant neurons still

66   systematically provide relevant information, and can serve as a substrate for PL despite variability.

67   However, these neurons are often not very precise: e.g., they signal which orientation is presented,

68   but cannot distinguish orientations as accurately as neurons in early visual areas[12].

69   Alternatively, the visual system could specialize on both task-relevant and -irrelevant aspects of the

70   stimuli, using highly precise neurons narrowly tuned to both dimensions (Fig. 1B). This strategy is

71   optimal when there is little variability because PL can be concentrated on the most informative

72   neurons. When facing high variability, this strategy implies that separate neurons need to be recruited

73   for each stimulus alternative during training. E.g., if oriented stimuli vary in spatial frequency (SF)

74   during orientation discrimination PL, each orientation-SF band combination needs to be learned

75   separately. This strategy readily assures that task performance relies on the most precise orientation

76   information available. Yet, by relying on narrowly tuned neurons, this strategy comes at a cost,

77   because these neurons cannot, by themselves, support generalization to other stimuli or locations in

78   the visual field.

79   Here, we test how the visual system deals with the challenge of variability while achieving high

80   performance in PL. We chose two features which are known to entail highly specific PL effects,

81   orientation and SF[13]. During training, we systematically vary the required precision of the orientation

82   discrimination task; in addition, subjects are trained with different degrees of variability in a task-

83   irrelevant dimension, SF. We find that task-irrelevant SF variability indeed leads to better

84   generalization of orientation discrimination performance to new SFs that were never shown during

85   training, even in difficult tasks. Furthermore, subjects trained with variable stimuli can generalize

86   better to new, untrained locations. Together, this suggests that they rely on SF-invariant neurons with

87   large receptive fields. We then perform the same experiments in a deep neural network (DNN) that

88   recapitulates several known PL phenomena[14]. We find a similar pattern of results, suggesting that

89   variability-induced catalysis of generalization holds in vivo as much as in silico. We go on to show that

90   SF variability during training leads to the recruitment (or emergence) of SF-invariant representations

91   – and not to an increase of SF-specialized units – highlighting the benefits of invariant representations
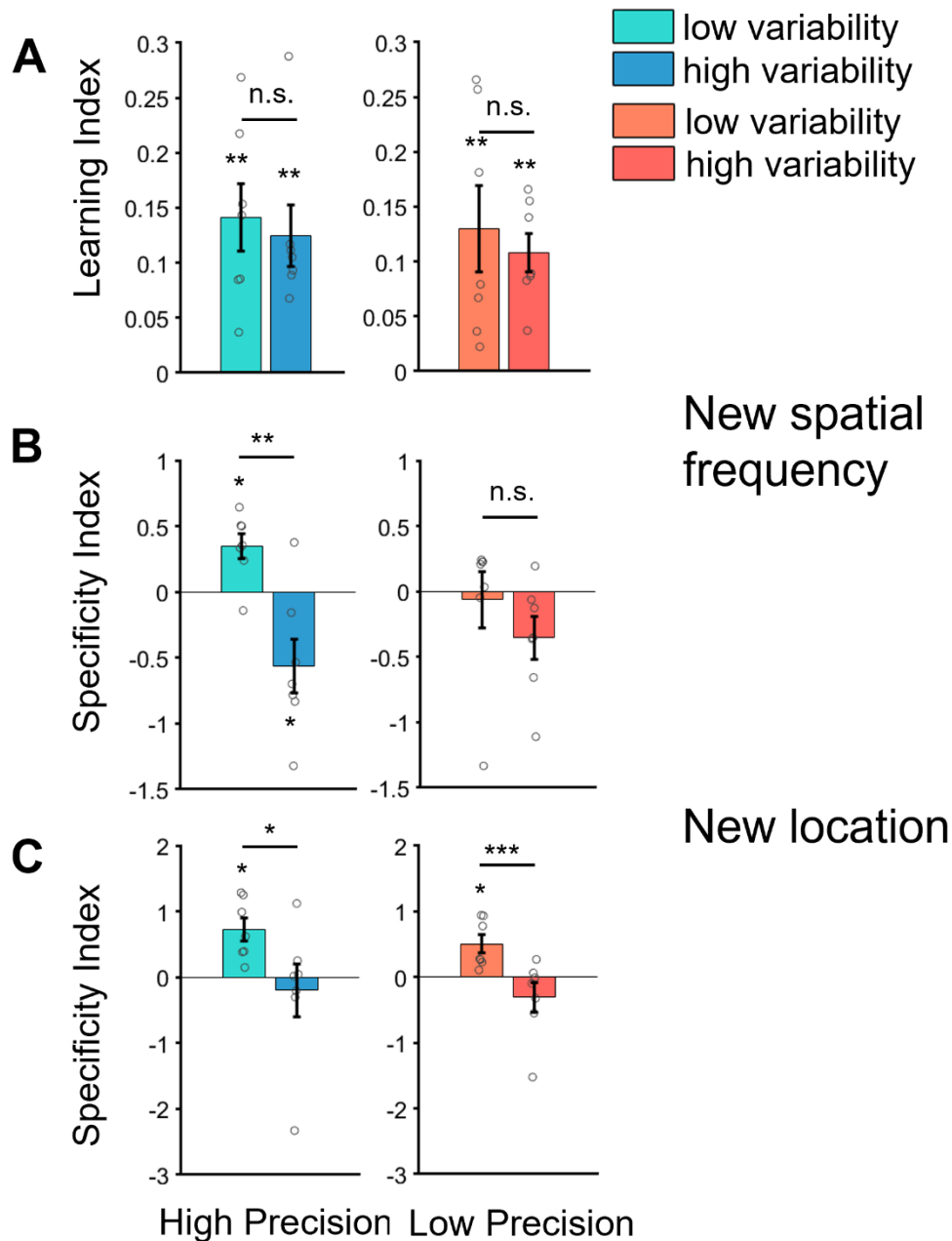
92   for generalization.

93

94   **Results**

95   We trained four groups of subjects (*n*=28) in an orientation discrimination task for several days, in

96   which they had to determine whether a grating was tilted clockwise or counterclockwise from a

97   reference on every trial. Two groups learned a high precision version of the task, with orientation

98   differences between 0.5 and 2.75 deg, while two other groups learned a low precision version of the

99   task, with orientation differences between 3 and 5.25 deg. One group from each precision level was

100 trained with a single SF (1.70 cpd), referred to as 'low variability' from here on, while the other group

101 received training with three pseudo-randomly interleaved SFs (0.53, 1.70 and 2.76 cpd), referred to

102 as 'high variability'.

103 All four groups showed statistically significant increases in orientation discrimination performance as

104 a function of training: on average, orientation thresholds decreased by 1.36° in the four training

105 groups (Supplemental Fig. S1). Expressed in units of "Learning Index" (LI)[15], which quantifies learning

106 relative to the baseline performance level, the average improvement was LI=0.13 (Fig. 2A). In fact,

107 there were no significant differences in LIs between the groups (precision main effect $F(1, 24)=0.32$,

108 $p=0.574$, partial $\eta^2=0.01$; variability main effect $F(1, 24)=0.03$, $p=0.863$, partial $\eta^2=0.02$; interaction

109 effect $F(1, 24)=0.05$, $p=0.830$, partial $\eta^2=0$). This suggests that variability did not negatively affect the

110 amount of PL of the trained task.

111

Figure 2: Learning and generalization in human subjects. A. Learning in orientation discrimination, quantified as Learning Index (LI), was significant in all groups (high precision groups; low variability mean LI=0.14, permutation test, *p*=0.008, Hedges' *g*=2.31; high variability mean LI=0.12, permutation test, *p*=0.008, Hedges' *g*=2.23; low precision groups; low variability mean LI=0.13, permutation test, *p*=0.008, Hedges' *g*=1.66; high variability mean LI=0.11, permutation test, *p*=0.008, Hedges' *g*=3.06). Task-irrelevant variability did not negatively affect the amount of learning (high precision groups mean difference in LI=0.017, exact permutation test, *p*=0.702, Hedges' *g*=0.22; low precision groups mean difference in LI=0.022, exact permutation test, *p*=0.615, Hedges' *g*=0.27). We obtained the same results when we quantified learning on the basis of orientation discrimination thresholds

122   (Supplemental Fig. S1) and when we trained with a different range of spatial frequencies (1.70, 2.54

123   and 2.76 cpd) in the high precision, high variability group (Supplemental Fig. S2). **B**. To test the

124   specificity of learning, a new SF was presented to the subjects after the last training session. After high

125   precision training, learning was specific for the group trained with low variability (mean SI=0.35,

126   permutation test, p=0.031, Hedges' g=1.81). In contrast, subjects that were trained on the same high

127   precision task but with varying SFs fully generalized to the new SF (mean SI=-0.56, permutation test,

128   p=0.047, Hedges' g=-1.38). Accordingly, there was a statistically significant difference between low

129   and high variability training within the high precision group (mean difference in SI=0.91, exact

130   permutation test, p=0.004, Hedges' g=2.16). Both low precision groups showed generalization (low

131   variability mean SI=-0.06, permutation test, p=0.969, Hedges' g=-0.15; high variability mean SI=-0.35,

132   permutation test, p=0.078, Hedges' g=-1.10), with no difference between them (mean difference in

133   SI=0.29, permutation test, p=0.317, Hedges' g=0.58). We obtained the same results when we

134   quantified specificity on the basis of orientation discrimination thresholds (Supplemental Fig. S1) and

135   after training with a different range of spatial frequencies (1.70, 2.54 and 2.76 cpd) in the high

136   precision, high variability group (Supplemental Fig. S2). **C**. We also tested whether subjects could

137   perform the transfer task at a new location. Here, training variability shaped specificity and

138   generalizability of learning similarly in the high and low precision regimes (low variability mean

139   difference in SI=0.92, exact permutation test, p=0.033, Hedges' g=1.14; high variability mean

140   difference in SI=0.81, permutation test, p<0.0001, Hedges' g=1.65): low variability groups both show

141   specificity (mean SI 0.72 and 0.50, respectively; permutation test, p=0.016, Hedges' g=2.12, and

142   p=0.016, Hedges' g=1.84). In contrast, training with high variability led to generalization (mean SI -

143   0.20 and -0.31, respectively; permutation test, p=0.703, Hedges' g=-0.25, and p=0.203, Hedges' g=-

144   0.69). We obtained the same results when we quantified specificity on the basis of orientation

145   discrimination thresholds (Supplemental Fig. S1). In all panels, *** stand for p<0.001, ** for p<0.01,

146   and * for p<0.05. Error bars reflect the standard error of the mean, circles reflect individual subjects.

147

148   We then tested generalization of PL in the four groups. For analyses, we computed the "Specificity

149   Index" (SI,[16]), which quantifies how much of the learning improvement can be carried over to

150   previously untrained conditions. Positive SI values indicate that learning is specific, i.e., does not

151   transfer to new conditions, whereas SI values smaller than or equal to 0 indicate generalization.

152   We first challenged subjects with a new, untrained SF (0.96 cpd) outside the SF channels of the closest

153   trained SFs. A 2 × 2 analysis of variance (ANOVA) of the four SIs revealed a statistically significant

154   interaction between the factors precision and variability (Fig. 2B, $F_{(1, 24)}$=4.52, p=0.044, partial

155   η^2=0.11): PL in the high precision, low variability group was highly specific to the trained SF (Fig. 2B

156   left, mean SI=0.35, permutation test, *p*=0.031, Hedges' *g*=1.81), as would be expected from a classical

157   PL training paradigm with only a single SF. In contrast, subjects that were trained on the same high

158   precision task but with varying SFs fully generalized to the new SF and in fact showed negative SI

159   (mean SI=-0.56, permutation test, *p*=0.047, Hedges' *g*=-1.38), implying that they even continued to

160   improve their performance in the transfer task. To rule out that generalization to 0.96 cpd was due to

161   the bracketing of this new SF by two trained SFs (0.53 and 1.7 cpd) in the high variability group,

162   compared to only one SF (1.7 cpd) in the low variability group, we additionally performed the same

163   experiment in a new group of subjects trained at 1.7, 2.54, and 2.76 cpd. Here, variability also led to

164   better generalization than low variability in the absence of bracketing (Supplemental Fig. S2B). The

165   results from high precision training thus suggest that task-irrelevant variability indeed enables

166   generalization in PL.

167   In the low precision group, a different picture emerged (Fig. 2B right). Here, both low and high

168   variability training led to good generalization with SIs not significantly different from 0 (low variability

169   mean SI=-0.06, permutation test, *p*=0.969, Hedges' *g*=-0.15; high variability mean SI=-0.35,

170   permutation test, *p*=0.078, Hedges' *g*=-1.10), and no statistically significant difference between both

171   training regimes (mean SI difference=0.29, exact permutation test, *p*=0.317, Hedges' *g*=0.58). This is

172   in accordance with previous PL studies showing that easy tasks generalize well[16] (but see below).

173   The results from the SF transfer condition suggest that high variability and low precision both enable

174   generalization. Given that subjects trained with high precision or low variability could perform the task

175   in a SF band that lay outside the trained SF range is indicative of a strategy involving SF-invariant

176   neurons. These neurons are more prevalent in higher order visual cortex, where neurons also have

177   larger receptive fields[17]. We thus hypothesized that if subjects relied on invariant neurons (and not

178   many narrowly tuned neurons which are more prevalent in early visual areas), they should also show

179   transfer to new spatial locations. This is because trained and untrained locations would be covered by

180   the same receptive fields. We thus moved the stimuli to a new, iso-eccentric location in the same

181   quadrant 8 dva from the original training location and repeated the transfer task there.

182   Here, we find that all groups trained with only one SF cannot transfer their learning gains to the new

183   location, irrespective of the required precision in the orientation discrimination task (mean SI 0.72 and

184   0.50, respectively; permutation test, *p*=0.016, Hedges' *g*=2.12, and *p*=0.016, Hedges' *g*=1.84). In

185   contrast, both groups that were trained with variable SFs were able to generalize (mean SI -0.20 and

186   -0.31, respectively; permutation test, *p*=0.703, Hedges' *g*=-0.25, and *p*=0.203, Hedges' *g*=-0.69). This

187   was also evident in terms of a main effect of variability in the ANOVA (Fig. 2C, *F*(1,24)=23.08, *p*<0.0001,

188   partial η^2=0.32; all other p>0.27, partial η^2<0.02). These results suggest that learning with variability

189   indeed taps on SF-invariant neurons with larger receptive fields than classical PL with only a single SF

190   band.

191   To gain further insight into the computations underlying generalization after high variability training,

192   we repeated the same experiments in a DNN purpose built for PL[14]. This network, which is derived

193   from the general AlexNet architecture[18], recapitulates several known behavioral and physiological PL

194   effects, e.g., higher learning rates for low precision tasks[16] and the sharpening of tuning curves as a

195   result of training that has originally been observed in the primary visual cortex of non-human

196   primates[19].

197
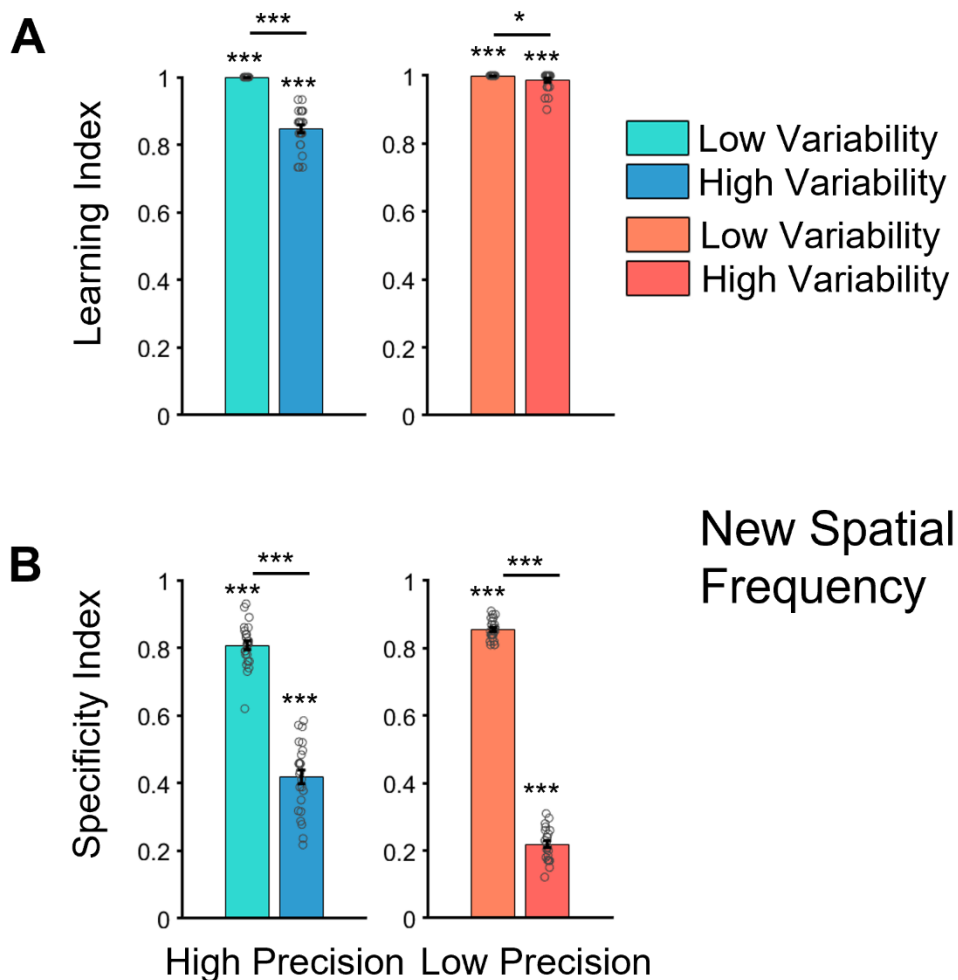
198

199   **Figure 3: Learning and generalization in deep neural networks. A**. We found significant learning

200   effects across all four training regimes with high learning levels (mean LI of 1, 0.85, 1 and 0.98

201   respectively, all p<0.0001, t(24)>14.01). Input variability led to lower LIs than training without

202   variability (high precision groups mean difference in LI=0.152, t(48)=12.97, p<0.0001, Hedges' g=3.61;

10

203 low precision groups mean difference in LI=0.013, $t(48)$=2.44, p=0.006, Hedges' g=0.69) **B**. When the

204 networks were fed with stimuli at a new untrained SF, we found specificity in all four groups (mean SI

205 0.81, 0.42, 0.85, and 0.22, respectively, all p<0.0001, $t(24)$>3.96). Importantly, the difference in

206 learning specificity was strongly affected by training variability (high precision groups mean difference

207 in SI=0.39, $t(48)$=16.03, p<0.0001, Hedges' g=4.46; low precision groups mean difference in SI=0.64,

208 $t(48)$=55.29, p<0.0001, Hedges' g=15.39). In all panels, *** stand for p<0.001, ** for p<0.01, and * for

209 p<0.05. Error bars reflect the standard error of the mean, circles reflect individual simulations.

210

211 We trained this network with our four training paradigms, crossing the factors precision and

212 variability, as in our human subjects. Each condition was simulated 25 times over 360 training steps.

213 When considering the final performance on the trained task, we again find statistically significant

214 learning effects in all four groups (Fig. 3A; mean LI 1, 0.85, 1, and 0.98, respectively, all p<0.0001, all

215 $t(24)$>72.36, all Hedges' g>14.01). Yet, in contrast to human subjects, there is a significant interaction

216 between variability and precision in LIs ($F(1,96)$=115.17, p<0.0001, partial η^2=0.54), with a larger

217 difference between high and low variability training in the high precision than the low precision

218 simulations (mean difference in LI=0.138, p<0.0001, Hedges' g=2.99). This difference notwithstanding,

219 critically, when we challenge the network with a new, untrained SF, we observe a differential pattern

220 of generalization performance that resembles what we found in human observers, most clearly in the

221 high precision training regime: networks trained with variable SFs show better transfer to a new SF

222 than networks trained with only a single SF (Fig. 3B; mean difference in SI=0.39, $t(48)$=16.03, p<0.0001,

223 Hedges' g=4.46). We obtain a similar result with low precision (mean difference in SI=0.64,

224 $t(48)$=55.29, p<0.0001, Hedges' g=15.39). Across training regimes, variability explained most between-

225 group variance (variability main effect $F(1, 96)$=1457.58, p<0.0001, partial η^2=0.94, compared to

226 precision main effect $F(1, 96)$=32.18, p<0.0001, partial η^2=0.25, and interaction $F(1, 96)$=85.01,

227 p<0.0001, partial η^2=0.47). Hence, while training leads to different LIs and more specific learning

228 effects in DNNs than in humans, the overall pattern of generalization results bears resemblance

229 between our in vivo and in silico results.

230 To dissociate a generalization strategy relying on SF-invariant orientation representations from a

231 specialization strategy relying on units narrowly tuned to SF and orientation, we then investigated

232 how representations in the DNN changed as a function of training. Since the amount of training

233 involved in the network simulations cannot easily be related to the number of trials seen by human

234 observers, we correlated the pattern of results that we obtained with humans to the pattern of

235 transfer results after 2, 5, 15, 35, 45, 50, 100, and 360 training steps of the networks. We find the

236 highest correspondence between in vivo and in silico experiments after 100 training steps (correlation

237 of SI between human subjects and DNNs, *r*=0.81, *p*<0.0001). We used the simulation results from this

238 training step for all subsequent analyses. To compare the two strategies, we computed the number of

239 units tuned to the trained SF and a SF-invariant orientation tuning index (SIOI) and for each of the 5

240 layers of the DNN (see STAR methods).

241
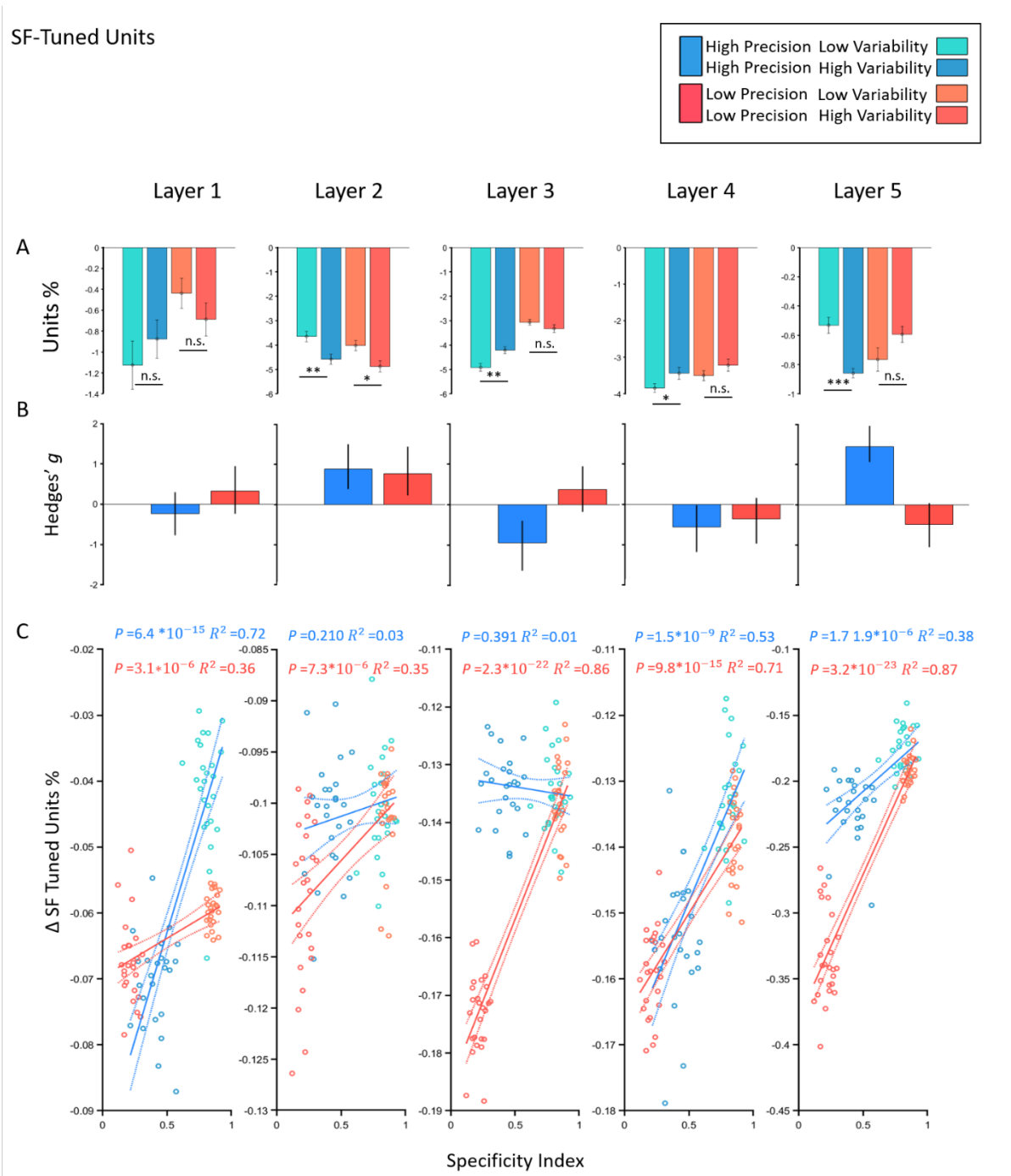
242

243 **Figure 4: Effects of learning on the number of SF-tuned units. A**. The number of SF-tuned units
244 decreases significantly across all layers and all groups (all $p<0.003$). Error bars reflect the standard
245 error of the mean. *** stand for $p<0.001$, ** for $p<0.01$, and * for $p<0.05$. **B**. The effect size for
246 comparing the change in the number of SF-tuned units between low and high variability groups
247 (expressed as Hedges' g) varies in size and direction across layers and training regimes (0.24 and 0.32
248 for layer 1; 0.88 and 0.77 for layer 2; -0.94 and 0.37 for layer 3; -0.56 and -0.36 for layer 4 and 1.44
249 and -0.49 for layer 5). Error bars reflect the 95% confidence interval of the effect size. **C**. The change
250 in the number of SF-tuned units is predictive of learning specificity in layer 1 ($r=0.85$, $p<0.0001$,
251 $R^2=0.72$), layer 4 ($r=0.73$, $p<0.0001$, $R^2=0.53$), and layer 5 ($r=0.62$, $p<0.0001$, $R^2=0.38$) in high precision
252 training (all other layers magnitude $r<0.02$, $p>0.210$, $R^2<0.03$), and in all layers in low precision training
253 (all $r>0.59$, $p<0.0001$, $R^2>0.34$). Dotted lines reflect 95% confidence intervals.

254

255 We find that instead of increasing, learning significantly decreases the number of SF-tuned across all
256 layers (Fig. 4A and Supplemental Table S1). Hence, a specialization strategy is unlikely to explain our
257 results. The decrease in the number of SF-tuned neurons differentiated high from low variability
258 training in several layers (Fig. 4B and Supplemental Table S1) and was predictive of learning specificity
259 (Fig. 4C) in layer 1 ($r=0.85$, $p<0.0001$, $R^2=0.72$), layer 4 ($r=0.73$, $p<0.0001$, $R^2=0.53$), and layer 5 ($r=0.62$,
260 $p<0.0001$, $R^2=0.38$) in high precision training (all other layers magnitude $r<0.02$, $p>0.210$, $R^2<0.03$),
261 and in all layers in low precision training (all $r>0.59$, $p<0.0001$, $R^2>0.34$). This provides a first hint at a
262 generalization strategy. We thus turned to SIOI to further investigate the relationship between
263 invariance and generalization. Indeed, we find that SF-invariant orientation tuning as quantified by
264 SIOI increases as a function of training, especially in the high variability groups, with the largest effects
265 in the top layer of the DNN (see Fig. 5A, B and Supplemental Table S2). Furthermore, from layer 1
266 onwards, SIOI predicts generalization on the transfer task (Fig. 5C) in high precision training (layer 1
267 $r=-0.30$, $p=0.035$, $R^2=0.09$; layer 2 $r=-0.88$, $p<0.0001$, $R^2=0.77$; layer 3 $r=-0.85$, $p<0.0001$, $R^2=0.72$; layer
268 4 $r=-0.86$, $p<0.0001$, $R^2=0.74$; layer 5 $r=-0.89$, $p<0.0001$, $R^2=0.78$) and low precision training (layer 1 $r=-$
269 0.78, $p<0.0001$, $R^2=0.61$; layer 2 $r=-0.95$, $p<0.0001$, $R^2=0.90$; layer 3 $r=-0.98$, $p<0.0001$, $R^2=0.96$; layer
270 4 $r=-0.98$, $p<0.0001$, $R^2=0.96$; layer 5 $r=-0.97$, $p<0.0001$, $R^2=0.97$). This suggests that variability-induced
271 generalization in PL could indeed rely on a recruitment (or emergence) of invariant representations
272 that provide orientation information irrespective of SF.

273
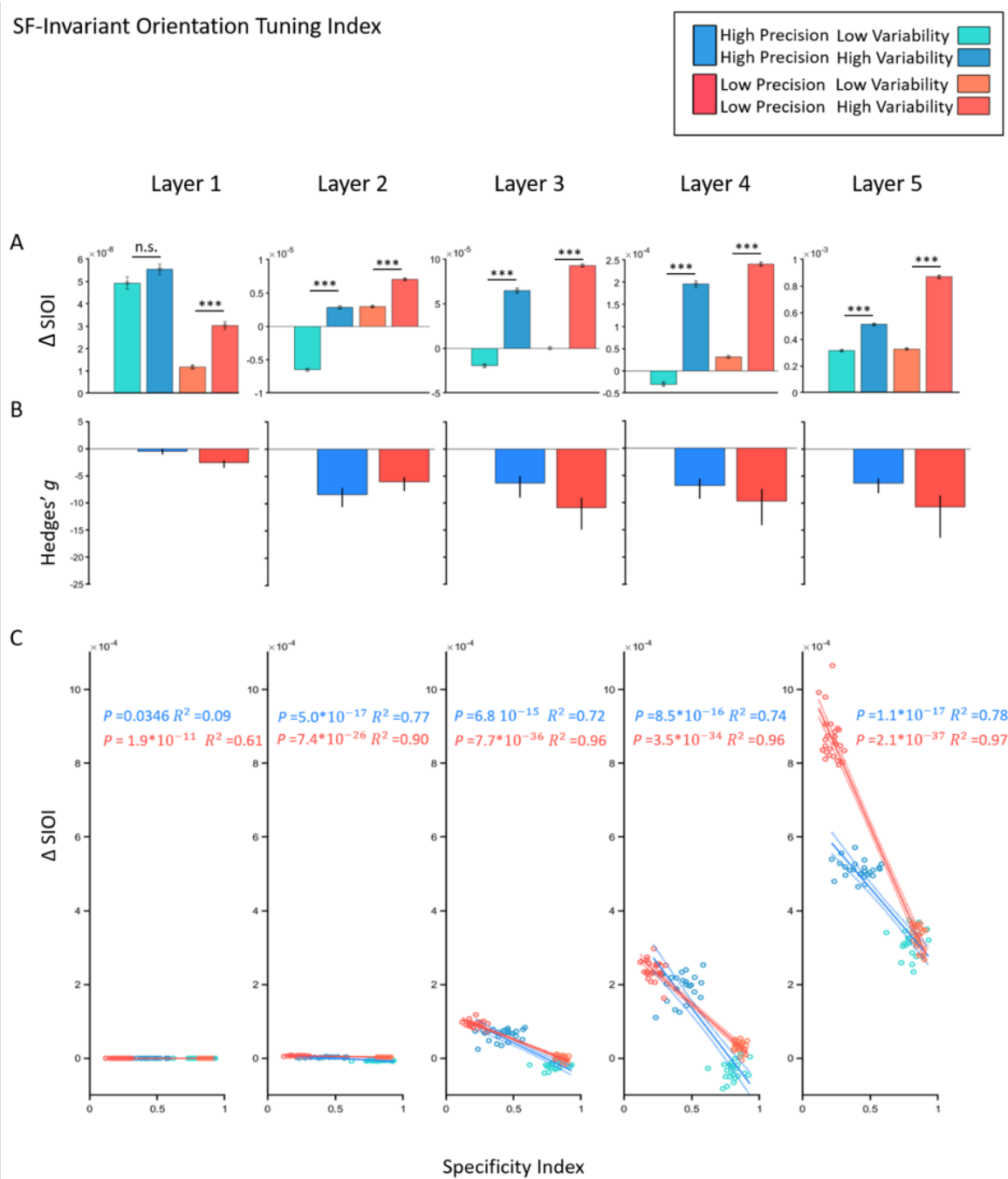
Figure 5: Effects of learning on SF-invariant orientation tuning. A. With low precision training, SIOI increases significantly (all $p<0.0001$) over training in all layers (with the exception of layer 3, where there is no change in SIOI for low variability training, $p=0.994$). This general trend is conserved for high precision, high variability training (all $p<0.0001$), while the low variability group shows significant SIOI decreases in layers 2, 3 and 4 (all $p<0.0001$). Error bars reflect the standard error of the mean. *** stand for $p<0.001$, ** for $p<0.01$, and * for $p<0.05$. **B**. The effect size for comparing the change in SIOI between low and high variability groups (expressed as Hedges' g) differentiates high from low variability training in all layers for high and low precision and increases across layers (-0.48 and -2.60

14

283 for layer 1; -8.33 and -6.00 for layer 2; -6.34 and -10.91 for layer 3; -6.83 and -9.70 for layer 4 and -

284 6.37 and -10.68 for layer 5). Error bars reflect the 95% confidence interval of the effect size. **C**. A strong

285 negative correlation between SIOI and SI is found from layer 1 onwards (all $p<0.035$) and increases

286 over layers. Dotted lines reflect 95% confidence intervals.

287 **Discussion**

288 We find that variability enables generalization in PL, beyond the generalization benefits that have been

289 reported in low precision tasks. Variability-induced generalization outside the trained SF band and far

290 from the trained location overcomes two of the hallmarks of PL, namely stimulus and location

291 specificity. Our results in human subjects align with the known tuning properties of SF-invariant

292 neurons in the primate visual system. A role of such neurons in generalization is further suggested by

293 our in silico results, which show that networks trained with variable stimuli evolve SF-invariant

294 representations of orientation with training.

295 Traditionally, PL protocols have not involved variability, training with a minimal number of stimuli

296 instead. This has led to many important insights, but mostly on extremely specific PL effects. E.g.,

297 simply showing stimuli to a different eye after monocular training can completely abolish learning

298 effects[20]. By eliminating variability and presenting the same stimuli over and over, traditional PL

299 paradigms may have unwittingly promoted 'memorization' or 'overfitting' of specific stimuli instead

300 of generalization[11]. Variability during training endows the visual system with more robustness towards

301 changes in the stimulus material. This also makes PL less "strange": the high specificity of PL has not

302 been observed in other domains of training – instead, variability has been shown to benefit transfer

303 in domains as disparate as baseball and soccer practice[21,22], language learning[23], and mathematics and

304 problem solving[24]. Hence, variability as a catalyst or enabler of generalization can be considered a

305 principle of leaning across domains.

306 Our data suggest that instead of memorizing a small number of specific stimuli, the visual system uses

307 invariant representations that provide orientation information irrespective of SF. Invariance is thought

308 to arise from systematic pooling over feature-specific representations. When there is variability in the

309 inputs, this entails more frequent weight updates for invariant than for feature-specific

310 representations, because only the former participate in the task on every trial. Furthermore, relying

311 on invariant representations suggests that fewer neurons need to undergo plasticity than in a

312 specialization strategy, hence reducing overall metabolic cost. Invariant representations for

313 orientation exist, e.g., in VTC, where neurons represent the orientation of simple stimuli like gratings

314 independently of other stimulus properties such as color or spatial frequency[25,26]. Although not as

315 frequently studied as orientation representations, e.g., in the primary visual cortex, these higher order

15

visual neurons have been shown to be causally relevant for orientation discrimination[27]. It has also been shown electrophysiologically that orientation tuning of neurons in higher order visual cortex is less accurate than in early visual areas[12]. This is one of the bases of the Reverse Hierarchy Theory (RHT)[16] that has previously put forward that easier tasks are learned on the basis of neurons in higher order visual cortex, whereas difficult tasks require high precision information available in early visual areas. Our results are in line with RHT, as subjects trained in the low precision task showed generalization to new SFs even if originally trained only with a single SF. However, variability during training led to overall lower SIs and extended generalization to new locations, even when the task was difficult. Hence, variability needs to be considered as an additional relevant dimension beyond task difficulty in enabling generalization of PL.

The pattern of SF and location transfer differed across the four training regimes in a way that may be informative regarding the physiological basis of the effects we observed. In particular, in the high precision groups, transfer to a new SF *and* to a new location only occurred under conditions of high variability. In the low precision group, SF transfer occurred in both the high and low variability groups, but location transfer was only evident after high variability training. This pattern of results suggests that visual PL could involve (at least) three populations of neurons depending on the training regime, namely: 1. neurons tuned to SF and location, as they are commonly found in early visual cortex; these neurons could explain the SF and location specificity in the high precision, low variability group. 2. neurons invariant to SF but not location, as they can be found at intermediate stages of visual processing; these neurons could explain concomitant transfer across SF and specificity for location in the low precision, low variability group. 3. neurons invariant to location and SF, which can be found in higher visual cortex, and which could explain transfer across SF and location in the high variability training groups across precision levels. Overall, this suggests that the contribution of lower versus higher visual areas in PL is not fixed but may depend on the variability of the training regime, which can be parsimoniously explained by a single principle, invariance (in space and SF, respectively). Future studies could test whether this also holds for other feature combinations.

Traditional PL theories have highlighted the specificity of PL[28], but more recent studies have identified conditions other than variability under which PL can generalize (for a recent review, see [29]). For example, in so-called 'double training' paradigms, practicing two tasks enables transfer across locations[30], and in 'training-plus-exposure' paradigms, simultaneous or subsequent passive exposure to additional stimuli enables transfer across orientations[31]. It has also been proposed that counteracting adaptation that may arise during prolonged training can be beneficial for generalization[32]. To this end, task-irrelevant 'dummy' trials can be interspersed with the main task, which is akin to introducing task-irrelevant variability. We did not observe statistically significant

350   adaptation effects in our data (Supplemental Fig. S4), possibly because we minimized stimulus-specific

351   adaptation by randomizing phase from trial to trial. However, our experiments were not specifically

352   designed to investigate the role of adaptation. To what extend the above-mentioned training

353   paradigms can be understood under the same principle(s) remains an interesting and very relevant

354   question.

355   Our results highlight (at least) three objectives for future studies: First, they suggest new targets for

356   electrophysiological recordings investigating PL in non-human primates, namely higher order visual

357   areas in VTC. Neurons in these areas code for orientation, are invariant to low-level features, and have

358   large spatial receptive fields, and could thus support the generalization of PL we observed. Yet, studies

359   investigating PL in these areas are, by and large, inexistent (but see[33]). Second, our results suggest that

360   stimulus variability-induced generalization is a robust principle of learning across many learning

361   domains, including vision, that may be strategically used in machine learning to achieve out of sample

362   generalization. Replacing ad hoc data augmentation strategies by biologically inspired principles may

363   lead to more robust models that may also align better with human perception. Finally, in terms of

364   application of VPL[34], where generalization is key, it may prove beneficial to counteract overtraining by

365   varying stimuli in task-irrelevant dimensions.

366

372

373   **CRediT author statement**

374   G.M., Methodology, Investigation, Formal Analysis, Visualization, Data curation, Writing – original

375   draft preparation; A.S.D., Software; C.M.S., Conceptualization, Methodology, Writing – original draft

376   preparation, Supervision, Project administration, Funding acquisition.

377

378   **Declaration of interests**

379   The funders had no role in study design, data collection and interpretation, decision to publish, or

380   preparation of the manuscript. ASD is a founder of Neuro-Inspired Vision and a member of its scientific

381   advisory board. GLM and CMS declare no competing financial interests.

382

383   **STAR METHODS**

384   **RESOURCE AVAILABILITY**

385   **Lead contact**

386   Further information and requests for resources and reagents should be directed to and will be fulfilled

387   by the lead contact, Caspar M. Schwiedrzik (c.schwiedrzik@eni-g.de).

388   **Materials availability**

389   This study did not generate new unique reagents.

390   **Data and code availability**

391   • All human behavioral data are available for download on (link available after journal
392   acceptance).
393   • The deep neural network is available from the original authors on Github
394   (https://github.com/kevin-w-li/DNN_for_VPL).
395   • Any additional information required to reanalyze the data reported in this paper is available
396   from the lead contact upon request.

397

398   **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

399   **Human Participants**

400   A total of 60 healthy human volunteers (34 female, 10 left-handed, mean age 28 yrs, SD 8.4 yrs)

401   participated in this study: 37 in the main experiments (20 female, 6 left-handed, mean age 28 yrs, SD

402   7.2 yrs), 7 in a pilot experiment for the range of SFs (3 female, 0 left-handed, mean age 30 yrs, SD 10.5

403   yrs), 7 in a control experiment for potential pre-existing differences at the training and transfer

404   location (5 female, 1 left-handed, mean age 26 yrs, SD 3.8 yrs), and 9 to test for an effect of differential

405   bracketing of SFs (6 female, 3 left-handed, mean age 31 yrs, SD 9 yrs). All subjects had normal or

406   corrected-to-normal vision, no neurological or psychiatric disease, and gave written informed consent

407   before participation in accordance with the Declaration of Helsinki. Sample size was selected based

408   on previous studies. Subjects were randomly assigned to one of four training groups (crossing the

18

409 factors precision (high/low) and variability (high/low), see below). All procedures were approved by

410 the Ethics Committee of the University Medical Center Göttingen (protocol number 29/8/17).

METHOD DETAILS

General setup

413 All subjects were trained on a two alternative forced choice (2AFC) orientation discrimination task

414 with oriented gratings. Training lasted 2 to 5 days with one training session per day. Transfer tasks

415 were conducted later depending on subject availability (mean interval 2.6 days). Total training and

416 transfer time was 4 to 7 days. Stimuli were presented on an LCD monitor (ViewPixx EEG, refresh rate

417 120 Hz, resolution 1920 × 1080 pixel, viewing distance 65 cm) in a darkened, sound-attenuating booth

418 (Desone Modular Acoustics). Subjects viewed the screen through an elliptical aperture that covered

419 the screen edges. Stimulus delivery and response collection were controlled using Psychtoolbox[35]

420 running in Matlab (The Mathworks, Inc.). Auditory feedback was delivered via headphones

421 (Sennheiser HDA 280). During all experiments, we continuously acquired pupil and gaze

422 measurements using a high-speed, video-based eye tracker (SR Research Eyelink 1000+). Data were

423 sampled at 1000 Hz from both eyes. Subjects were paid €8 per hour. To assure constant motivation

424 over the training sessions, subjects received a bonus of €2 if they improved by 10% from the previous

425 training session.

Stimuli and task

427 On each trial, subjects had to decide whether a monopolar, monochromatic Gabor grating (size 3.1

428 dva, luminance 43.4 cd/m^2) was tilted clockwise or counterclockwise with respect to a reference

429 stimulus. The reference stimulus was a monopolar, monochromatic Gabor grating with identical size

430 and constant spatial frequency (SF, 2.56 cpd) which was randomly tilted per subject, avoiding

431 meridians and diagonals. The task stimuli were presented in ten linearly spaced difficulty levels

432 clockwise and counterclockwise from the reference, respectively. Each condition was presented 21

433 times. In addition, we presented the reference orientation 21 times, amounting to a total of 441 trials

434 per session evenly distributed among four blocks. In the high precision training groups, difficulty levels

435 ranged from 0.5 to 2.75 deg, while in the low precision groups, orientation differences ranged from 3

436 to 5.25 deg. In the low variability groups, all stimuli were presented at a single SF (1.70 cpd). In the

437 high variability groups, we used three SFs (0.53, 1.70 and 2.76 cpd). We chose these SFs such that they

438 lie outside the other spatial frequency channels (including the one used for transfer, see below), or

439 exactly at full width half maximum (FWHM), with the exception of 1.7 cpd which just falls into the 2.76

440 cpd band. For this, we assumed a SF channel bandwidth of 1.4 octaves[36] (but see [37]). In all groups,

19

441   stimuli were presented in pseudo-random order at 12.4° eccentricity against a grey background (39.5

442   cd/m^2). Phase varied randomly between 0° and 360° from trial to trial.

443   On each trial, we first presented the reference stimulus for 2000 ms. This was followed by a 1000 ms

444   fixation period. Then, two grey saccade placeholders (luminance 23.54 cd/m^2) appeared for 1500 ms,

445   followed by the stimulus for 250 ms. After a random delay of 500-3000 ms, the choice phase started,

446   which was indicated by a color change of the placeholders. To respond, subjects had to direct their

447   gaze from the fixation point (size 0.31 dva, luminance 0.28 cd/m^2) at the center of the screen to a red

448   target (luminance 23.53 cd/m^2) if the stimulus was rotated counterclockwise or to an isoluminant

449   green target if it was rotated clockwise. Note that although SF and orientation are not independently

450   processed in the visual system, we rendered SF task-irrelevant by instructing subjects to consider only

451   orientation for their task. The assignment of colors to the placeholder locations was pseudo-

452   randomized. Subjects had to reach the target at 12 dva distance within 1000 ms. Subjects were

453   instructed to respond as accurately as possible. Feedback on accuracy was provided by playing a low

454   pitch sound (incorrect) or a high pitch sound (correct) for 500 ms. The next trial started between 300

455   and 2000 ms later. If subjects did not respond in time or if they broke fixation (fixation window size

456   4.5 dva), the low pitch sound was played, and the trial was repeated later during the block. To keep

457   subjects motivated, the high pitch sound increased in loudness after the first 2 and 3 sequential correct

458   trials. Loudness was reset to the original level at the first incorrect trial.

459   Subjects were free to take breaks between blocks. There were additional breaks in the middle of each

460   block to display feedback about performance. As a reminder, the reference was shown for 2000 ms at

461   the center of the screen before and in the middle of each block (in total 8 times per session). Before

462   each training session, subjects performed 16 warm-up trials.

463   **Transfer conditions**

464   Subjects had to reach at least 0.05 Learning Index (see below) improvement and show stable

465   performance compared to the end of the preceding training session before we transitioned to the first

466   transfer session. Subjects generally required between 2 and 5 days to reach this criterion, and there

467   were no statistically significant differences in the amount of training days to reach criterion between

468   the groups ($F(3,24)=0.71$, $p=0.553$, partial $\eta^2=0.21$). For the first transfer test, we asked all subjects to

469   perform the same task as during training, but we changed the SF of the task stimuli to a new, unseen

470   SF (0.96 cpd). This SF was chosen to lie outside the FWHM of SF channels around the closest SFs used

471   for training (0.53 cpd and 1.7 cpd), assuming a channel bandwidth of 1.4 cpd[36]. Pilot experiments

472   showed that baseline performance at 0.96 cpd did not differ from baseline performance in the SFs

473   used for training (Supplemental Fig. S3). Subjects performed 441 trials within a single transfer session.

474 The SF of the reference remained identical. The interval between the last training session and the

475 transfer session varied between subjects (mean 2.6 days), but there were no significant differences

476 between the groups (F(3,24)=0.1, p=0.957, partial η^2=0.01). Furthermore, there was no significant

477 correlation between Specificity Indices (see below) and this interval (Pearson correlation, r=0.1031,

478 p=0.6015). For the second transfer test, we changed the location of the transfer stimuli to a new, iso-

479 eccentric position in the same quadrant 8 deg away from the original training location. Pilot

480 experiments showed that there were no systematic preexisting differences in orientation

481 discrimination performance between the training and the transfer location (mean difference in

482 accuracy 2%, permutation test, p=0.28, Hedges' g=0.21). Subjects again performed 441 trials within a

483 single session.

484

485 **Deep Neural Network Simulations**

486 The deep learning model used in this paper was adopted from[14]. The model was implemented in

487 PyTorch (version 1.10.0) and consists of two parallel streams, each encompassing the first five

488 convolutional layers of AlexNet[18] plus one fully connected layer which gives out a single scalar value.

489 One stream accepts one fixed reference stimulus and the other stream accepts one varying target

490 stimulus. The target stimulus is then compared to the reference stimulus. After the fully connected

491 layers, the outputs of the two parallel streams – two scalar values - are entered to a softmax layer to

492 give out one binary value which indicates the relative orientation (clockwise or counter-clockwise) of

493 the target stimulus relative to the reference stimulus. We used the same feature maps and kernel size

494 as the original paper. Each stimulus had 32×32 pixel size and was centered inside a 256×256 pixel

495 image which was cropped to 224×224 pixel shape to make it consistent with the input size of AlexNet.

496 All stimuli had homogenous gray background. The reference stimulus had a fixed orientation of 236°

497 and a SF of 2.56 cpd. While keeping the same orientation and spatial frequency combination as for

498 human subjects, we simulated four different training regimes, crossing the factors variability and

499 precision. Each regime was independently trained 25 times. We initialized the five convolutional layers

500 with pretrained ImageNet weights of AlexNet to mimic a (pretrained) adult brain. The last fully

501 connected layer was initialized by zero. As for human subjects, we randomized stimulus phase for

502 training. The training samples per group were set to batch sizes of 20 and 60 stimuli, for low variability

503 groups and high variability groups, respectively. This ensures that each batch provides all 20

504 orientations and all SFs but random stimulus phases. By randomizing stimulus phase, we created 7200

505 and 21600 unique stimuli for the low variability groups and for high variability groups, respectively.

506 Training parameters were set as follow: learning rate = 0.00001, momentum = 0.9, weight decay =

507 0.0001. The cross-entropy loss function was used as an objective function and optimized via stochastic

508 gradient descent.

509 **QUANTIFICATION AND STATISTICAL ANALYSIS**

510 All data analyses were carried out in Matlab (The Mathworks, Inc.) and R (version 4.2.1, R Core Team,

511 https://www.R-project.org).

512 **Human behavior**

513 Human behavioral data were analyzed using (exact) permutation *t*-tests (two-sided for comparisons

514 between groups, and one-sample tests against 0) and analyses of variance (ANOVA). Before fitting

515 ANOVAs, data were aligned and rank-transformed[38,39] using the ARTool package (version 0.11.1,

516 https://github.com/mjskay/ARTool) to satisfy distributional assumptions. For *t*-tests and ANOVAs, we

517 computed Hedges' *g* and partial η^2, respectively, as effect sizes. From the original sample of 37

518 subjects in the main experiments, 2 subjects did not complete the experiments and were thus

519 excluded from data analysis. 3 subjects were excluded from further data acquisition after the first

520 session because they evidently did not follow task instructions. 4 subjects were excluded during data

521 acquisition because lack of significant learning (but there was no difference in the number of excluded

522 subjects between high and low variability training regimes, *p*=0.6029, odds ratio = 3, Fisher exact test).

523 The final *n* in the main experiment was thus 28 (15 female, 5 left-handed, mean age 28 yrs, SD 7.6 yrs).

524 Accuracy was defined as the average percentage correct per session. We excluded all trials with

525 outliers in the reaction times per subject using the estimator Sn^{40} at a threshold of 8.5. To quantify

526 learning, we computed the Learning Index (LI)[15], as follows:

527
$$LI = \frac{Accuracy_{last\ trained} - Accuracy_{baseline}}{Accuracy_{baseline}}$$

528 where baseline accuracy was obtained by averaging the performance of the first session. Because LI

529 could not be lower than 0, we used one-sided (exact) permutation *t*-tests for comparisons against 0.

530 To quantify transfer, a Specificity Index (SI)[16] was calculated, as follows:

531
$$SI = \frac{Accuracy_{last\ trained} - Accuracy_{transfer}}{Accuracy_{last\ trained} - Accuracy_{baseline}}$$

532 In addition, we fitted psychometric functions (Weibull) using the Palamedes Toolbox[41] to derive

533 orientation discrimination thresholds per subject and session.

534 **Deep Neural Networks**

22

535 For each simulation, learning and transfer performances were quantified by computing LI and SI as

536 specified above. Baseline accuracy was obtained after the first training session. To investigate how the

537 different training conditions affected stimulus representations in the network, the trained networks

538 were evaluated with a set of stimuli covering all 20 orientations and all three SFs. We then computed

539 dissimilarities between the average channel activities across all pairwise stimulus combinations using

540 abs(1-Pearson correlation) for each channel pre and post training, respectively. Channels were

541 identified as SF-sensitive if the sum of the lower triangular matrix was bigger than zero. Furthermore,

542 we used the average channel activity to generate a representational similarity matrix (RSM) for each

543 of the five AlexNet layers. To quantify SF-invariant orientation tuning, we computed the SF-invariant

544 orientation tuning index (SIOI)[42,43] by dividing the mean correlation along the off-diagonals of this RSM

545 by the mean correlation of all other elements (but the main diagonal). To relate the pattern of

546 generalization results in humans to that of the deep neural networks, we first computed

547 representational dissimilarity of SIs separately for humans and deep neural networks. This step

548 resulted in a 4x4 representational dissimilarity matrix for humans and networks, respectively. In these

549 matrices, each dimension represents a given training regime, providing us with how dissimilar SIs were

550 in the four training regimes. Then, we compared the representational geometries of SIs between

551 humans and networks using Pearson correlation. All statistical analyses were carried out using

552 parametric *t*-tests (one-sided for LI and changes in SIOI, two-sided for all others), ANOVA, and Pearson

553 correlation.

554

555 **References**

556 1. Gold, J.I., and Stocker, A.A. (2017). Visual decision-making in an uncertain and dynamic world.

557 Annual review of vision science *3*, 227-250.

558 2. Herault, J., Guyader, N., and Guerin-Dugue, A. (2009). Scene variability and perception

559 constancy in the visual system: a model of pre-processing before data analysis and learning.

560 IEEE International Workshop on Machine Learning for Signal Processing, 1-12.

561 3. Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., and Sagi, D. (2004). Perceptual learning in

562 contrast discrimination: the effect of contrast uncertainty. Journal of vision *4*, 993-1005.

563 4. Yu, C., Klein, S.A., and Levi, D.M. (2004). Perceptual learning in contrast discrimination and the

564 (minimal) role of context. Journal of vision *4*, 169-182.

565 5. Herzog, M.H., Aberg, K.C., Fremaux, N., Gerstner, W., and Sprekeler, H. (2012). Perceptual

566 learning, roving and the unsupervised bias. Vision research *61*, 95-99.

567   6.   Zhaoping, L., Herzog, M.H., and Dayan, P. (2003). Nonlinear ideal observation and recurrent
568   preprocessing in perceptual learning. Network *14*, 233-247.

569   7.   Raviv, L., Lupyan, G., and Green, S.C. (2022). How variability shapes learning and
570   generalization. Trends in cognitive sciences *26*, 462-483.

571   8.   Patarnello, S., and Carnevali, P. (1987). Learning networks of neurons with boolean logic.
572   Europhysics letters *4*, 503-508.

573   9.   Shepard, R.N. (1987). Toward a universal law of generalization for psychological science.
574   Science *237*, 1317-1323.

575   10. Fahle, M., and Poggio, T., eds. (2002). Perceptual learning (MIT Press).

576   11. Sagi, D. (2011). Perceptual learning in vision research. Vision research *51*, 1552-1566.

577   12. Vogels, R., and Orban, G.A. (1994). Activity of inferior temporal neurons during orientation
578   discrimination with successively presented gratings. Journal of neurophysiology *71*, 1428-
579   1451.

580   13. Fiorentini, A., and Berardi, N. (1980). Perceptual learning specific for orientation and spatial
581   frequency. Nature *287*, 43-44.

582   14. Wenliang, L.K., and Seitz, A.R. (2018). Deep neural networks for modeling visual perceptual
583   learning. Journal of neuroscience *38*, 6028-6044.

584   15. Petrov, A.A., Van Horn, N.M., and Ratcliff, R. (2011). Dissociable perceptual-learning
585   mechanisms revealed by diffusion-model analysis. Psychonomic Bulletin & Review *18*, 490-
586   497.

587   16. Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning.
588   Nature *387*, 401-406.

589   17. Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual review of neuroscience *19*,
590   109-139.

591   18. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep
592   convolutional neural networks. Advances in neural information processing systems *25*.

593   19. Schoups, A.A., Vogels, R., Qian, N., and Orban, G.A. (2001). Practising orientation identification
594   improves orientation coding in V1 neurons. Nature *412*, 549-553.

595   20. Karni, A., and Sagi, D. (1991). Where practice makes perfect in texture discrimination:
596   evidence for primary visual cortex plasticity. Proceedings of the national academy of sciences
597   of the united states of america *88*, 4966-4970.

598   21. Hall, K.G., Domingues, D.A., and Cavazos, R. (1994). Contextual interference effects with
599   skilled baseball players. Perceptual and motor skills *78*, 835-841.

600   22.   Zetou, E., Papadakis, L., Vernadakis, N., Derri, V., Bebetsos, E., and Filippou, F. (2014). The
601   effect of variable and stable practice on performance and learning the header skill of young
602   athletes in soccer. Procedia social and behavioral sciences *152*, 824-829.

603   23.   Clopper, C.G., and Pisoni, D.B. (2004). Effects of talker variability on perceptual learning of
604   dialects. Language and speech *47*, 207-239.

605   24.   Van Merrienboer, J.J.G., and Kirschner, P.A. (2007). Ten steps to complex learning (Taylor &
606   Francis).

607   25.   Desimone, R., Schein, S.J., Moran, J., and Ungerleider, L.G. (1985). Contour, color and shape
608   analysis beyond the striate cortex. Vision research *25*, 441-452.

609   26.   Sary, G., Vogels, R., and Orban, G.A. (1993). Cue-invariant shape selectivity of macaque inferior
610   temporal neurons. Science *260*, 995-997.

611   27.   Adab, H.Z., and Vogels, R. (2016). Perturbation of posterior inferior temporal cortical activity
612   impairs coarse orientation discrimination. Cerebral cortex *26*, 3814-3827.

613   28.   Fahle, M. (2004). Perceptual learning: a case for early selection. Journal of vision *4*, 879-890.

614   29.   Lu, Z.-L., and Dosher, B.A. (2022). Current directions in visual perceptual learning. Nature
615   Reviews Psychology *1*, 654-668.

616   30.   Xiao, L.Q., Zhang, J.Y., Wang, R., Klein, S.A., Levi, D.M., and Yu, C. (2008). Complete transfer of
617   perceptual learning across retinal locations enabled by double training. Current biology *18*,
618   1922-1926.

619   31.   Zhang, J.Y., Zhang, G.L., Xiao, L.Q., Klein, S.A., Levi, D.M., and Yu, C. (2010). Rule-based learning
620   explains visual perceptual learning and its specificity and transfer. Journal of neuroscience *30*,
621   12323-12328.

622   32.   Harris, H., Gliksberg, M., and Sagi, D. (2012). Generalized perceptual learning in the absence
623   of sensory adaptation. Current biology *22*, 1813-1817.

624   33.   Adab, H.Z., Popivanov, I.D., Vanduffel, W., and Vogels, R. (2014). Perceptual learning of simple
625   stimuli modifies stimulus representations in posterior inferior temporal cortex. Journal of
626   cognitive neuroscience *26*, 2187-2200.

627   34.   Deveau, J., and Seitz, A.R. (2014). Applying perceptual learning to achieve practical changes in
628   vision. Frontiers in psychology *5*, 1166.

629   35.   Brainard, D.H. (1997). The Psychophysics Toolbox. Spatial vision *10*, 433-436.

630   36.   De Valois, R.L., and De Valois, K.K. (1990). Spatial vision (Oxford University Press).

631   37.   Guan, S.C., Ju, N.S., Tao, L., Tang, S.M., and Yu, C. (2021). Functional organization of spatial
632   frequency tuning in macaque V1 revealed with two-photon calcium imaging. Progress in
633   neurobiology *205*, 102120.

634 38. Higgins, J.J., and Tashtoush, S. (1994). An aligned rank transform test for interaction. Nonlinear
635   World *1*, 201-211.
636 39. Wobbrock, J.O., Findlater, L., Gergle, D., and Higgins, J.J. (2011). The Aligned Rank Transform
637   for Nonparametric Factorial Analyses Using Only Anova Procedures. Proceedings of the SIGCHI
638   Conference on Human Factors in Computing Systems, 143–146.
639 40. Rousseeuw, P.J., and Croux, C. (1993). Alternatives to the median absolute deviation. Journal
640   of the American Statistical Association *88*, 1273-1283.
641 41. Prins, N., and Kingdom, F.A.A. (2018). Applying the Model-Comparison Approach to Test
642   Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox.
643   Frontiers in psychology *9*, 1250.
644 42. Freiwald, W.A., and Tsao, D.Y. (2010). Functional compartmentalization and viewpoint
645   generalization within the macaque face-processing system. Science *330*, 845-851.
646 43. Farzmahdi, A., Rajaei, K., Ghodrati, M., Ebrahimpour, R., and Khaligh-Razavi, S.M. (2016). A
647   specialized face-processing model inspired by the organization of monkey face patches
648   explains several face-specific phenomena observed in humans. Scientific reports *6*, 25025.