Two-stage perceptual learning to break visual crowding

Ziyun Zhu*

Department of Psychology and Beijing Key Laboratory of
Behavior and Mental Health
Key Laboratory of Machine Perception (Ministry of Education)
Peking-Tsinghua Center for Life Sciences
PKU-IDG/McGovern Institute for Brain Research,
Peking University, Beijing, P.R. China ✉

Zhenzhi Fan*

Department of Psychology and Beijing Key Laboratory of
Behavior and Mental Health
Key Laboratory of Machine Perception (Ministry of Education)
Peking-Tsinghua Center for Life Sciences
PKU-IDG/McGovern Institute for Brain Research,
Peking University, Beijing, P.R. China ✉

Fang Fang

Department of Psychology and Beijing Key Laboratory of
Behavior and Mental Health
Key Laboratory of Machine Perception (Ministry of Education)
Peking-Tsinghua Center for Life Sciences
PKU-IDG/McGovern Institute for Brain Research,
Peking University, Beijing, P.R. China 🏠 ✉

When a target is presented with nearby flankers in the peripheral visual field, it becomes harder to identify, which is referred to as *crowding*. Crowding sets a fundamental limit of object recognition in peripheral vision, preventing us from fully appreciating cluttered visual scenes. We trained adult human subjects on a crowded orientation discrimination task and investigated whether crowding could be completely eliminated by training. We discovered a two-stage learning process with this training task. In the early stage, when the target and flankers were separated beyond a certain distance, subjects acquired a relatively general ability to break crowding, as evidenced by the fact that the breaking of crowding could transfer to another crowded orientation, even a crowded motion stimulus, although the transfer to the opposite visual hemi-field was weak. In the late stage, like many classical perceptual learning effects, subjects' performance gradually improved and showed specificity to the trained orientation. We also found that, when the target and flankers were spaced too finely, training could only reduce, rather than completely eliminate, the crowding effect. This two-stage learning process illustrates a learning strategy for our brain to deal with the notoriously difficult problem of identifying peripheral objects in clutter. The brain first learned to solve the "easy and general" part of the problem (i.e., improving the processing resolution and segmenting the target and flankers) and then tackle the "difficult and specific" part (i.e., refining the representation of the target).

Introduction

In peripheral vision, one's ability to identify a target is impeded by nearby flankers. This phenomenon is known as *crowding* (Levi, 2008; Whitney & Levi, 2011). Crowding has been reported to occur with many kinds of stimuli and tasks, such as letter recognition (Bouma, 1970), vernier acuity (Levi, Klein, & Aitsebaomo, 1985), orientation discrimination (Westheimer, Shimamura, & McKee, 1976), stereo acuity (Butler & Westheimer, 1978), and face recognition (Louie, Bressler, & Whitney, 2007). It sets a fundamental limit on visual perception and conscious awareness in the periphery.

Many theories have been advanced to explain visual crowding (Levi, 2008). Most of the theories

posit that at some stage of peripheral information processing, the visual system lacks the necessary resolution (constrained by retinal/cortical sampling and/or spatial attention) to process the target individually when it is surrounded by nearby flankers. Because of the limited resolution, features from the target and flankers are mistakenly integrated, resulting in a nonveridical percept. However, it is still highly controversial where the resolution bottleneck is in the visual processing stream. The two most popular theories for crowding have placed the bottleneck at the level of bottom-up feature pooling (Levi, Hariharan, & Klein, 2002; Pelli & Tillman, 2008) or attentional selection (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Strasburger, Harvey, & Rentschler, 1991). Although both theories have received empirical support, neither provides an adequate explanation for the large body of existent psychophysical and brain imaging data (Herzog & Manassi, 2015; Levi, 2008; Whitney & Levi, 2011).

Training can improve performance for many visual tasks, which is referred to as *visual perceptual learning* (Sagi, 2011; Watanabe & Sasaki, 2015). One might ask whether training could reduce crowding and improve peripheral vision. Recent studies (Chung, 2007; Huckauf & Nazir, 2007; Hussain, Webb, Astle, & McGraw, 2012; Sun, Chung, & Tjan, 2010; Xiong, Yu, & Zhang, 2015) have demonstrated that, following training, the accuracy of crowded letter identification and the spatial extent of crowding could be significantly reduced. This finding not only is theoretically interesting because it opens a new window to understand the mechanisms of visual crowding from a perspective of perceptual learning but also provides a new noninvasive treatment for children and adults with amblyopia.

Although previous studies have demonstrated a training-induced reduction of crowding, it is still unknown whether training can break or completely remove crowding. Here, in a series of psychophysical experiments, human subjects were trained on a crowded orientation discrimination task. We attempted to address two issues: (a) Under what conditions can crowding be completely eliminated by training? and (b) Can the elimination of crowding be transferred to other stimuli and locations? If crowding is determined by the processing resolution of the visual system and the resolution can be improved by training, we predict that crowding can be reduced, and even completely eliminated, by training if the target and flankers are separated by a certain distance. If the processing resolution is determined by high-level attentional selection, the breaking of crowding might be able to transfer to other stimuli.

Materials and methods

Subjects

There were eight subjects (four male) in Experiment 1, eight (four male) in Experiment 2, nine (four male) in Experiment 3, eight (four male) in Experiment 4, and 10 (six male) in Experiment 5. All subjects were right handed and reported normal or corrected-to-normal vision. None of them participated in more than one experiment. Ages ranged from 18 to 26 years. They gave written, informed consent in accordance with the procedures and protocols approved by the human subject review committee at Peking University. This study adhered to the Declaration of Helsinki.

Stimuli and design

Visual stimuli were displayed on an IIYAMA color graphic monitor (model MM906UT; refresh rate: 100 Hz; resolution: 1024×768; size: 19 in.) with a gray background (luminance: 47.59 cd/m^2). Subjects viewed the stimuli from a distance of 57 cm, and they were asked to maintain fixation on a black dot at the center of the display. Their head position was stabilized using a head and chin rest. Eye positions were not monitored in this study.

Experiment 1 consisted of five phases: pretraining test (Pre), orientation discrimination training (Training 1), mid-training test (Mid), orientation discrimination training (Training 2), and posttraining test (Post; Figure 1A). During the training phases, subjects were trained to perform an orientation discrimination task with a (crowded) target grating (radius: 1.5°; spatial frequency: 2 cycles/°; Michelson contrast: 1; mean luminance: 47.59 cd/ m^2; eccentricity: 10°; position: the upper-left visual quadrant) at an orientation of θ (Figure 1B). For each subject, the trained orientation θ was chosen randomly from two perpendicular orientations (67.5° and 157.5°; 0° was the horizontal) at the beginning of the experiment and was then fixed throughout the training phases. A daily training session (about 1 hr) consisted of 30 QUEST staircases of 40 trials (Watson & Pelli, 1983). In a trial, two (crowded) targets with orientations of θ and $\theta \pm \Delta\theta$ were presented successively for 200 ms each and were separated by a 600-ms blank interval (Figure 1B). The temporal order of these two targets was randomized. Two flankers were always positioned abutting the target in the radial direction with respect to fixation. The flankers were identical to the target except that their orientations were randomized. The center-to-center distance between the target and the flankers was 3°. So the flankers and target were abutting. Subjects
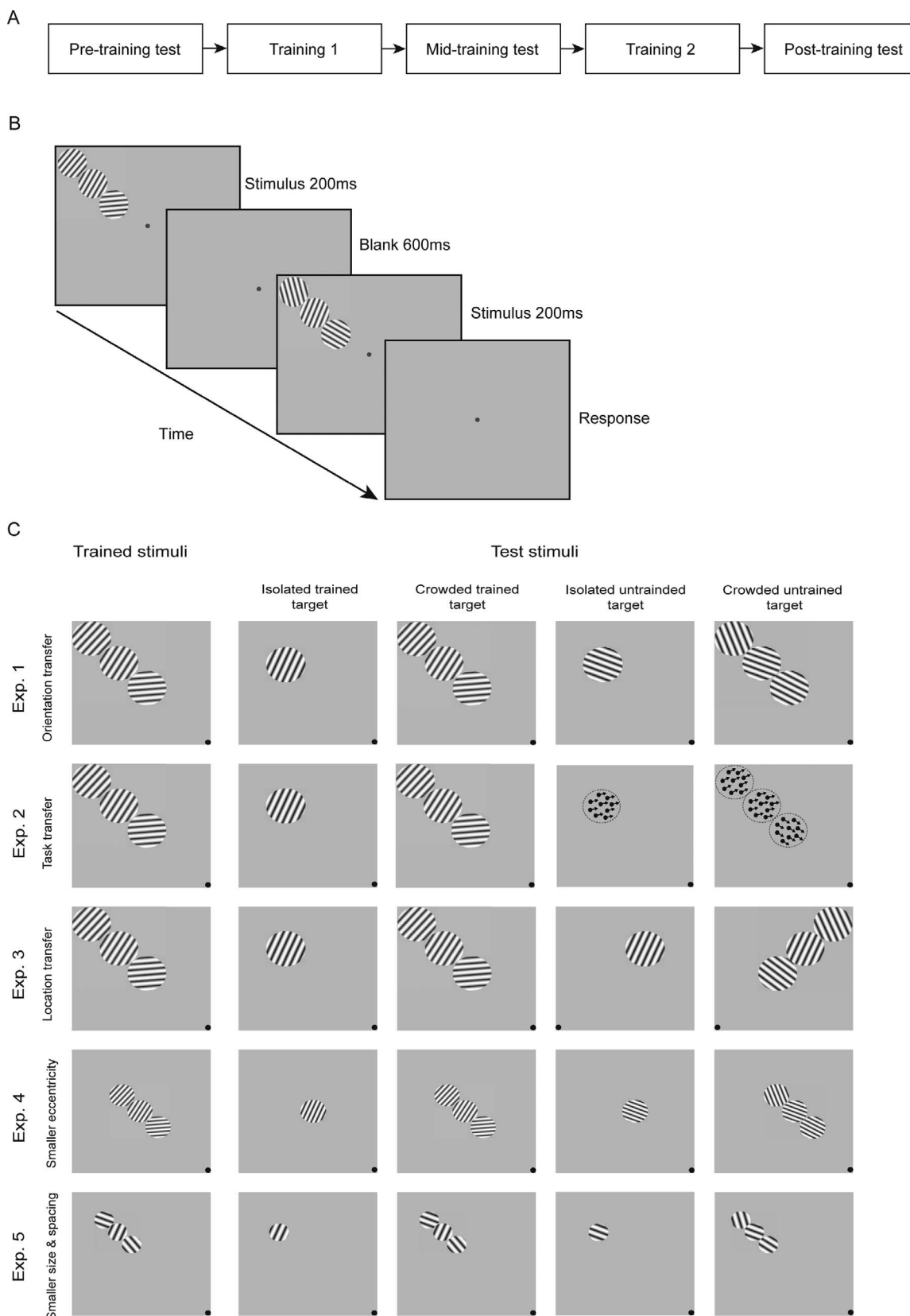
Figure 1. Experimental protocol and stimuli. (A) Each experiment consisted of five phases: pretraining test (Pre), Training 1, mid-training test (Mid), Training 2, and posttraining test (Post). (B) Schematic description of a two-alternative forced-choice trial in a QUEST staircase for measuring the orientation discrimination threshold with a crowded target. (C) Trained and test stimuli in Experiments 1–5. Black dots represent the fixation point. The stimuli were presented in the upper-left visual quadrant, except that the isolated and crowded untrained targets in Experiment 3 were presented in the upper-right visual quadrant.

were asked to make a two-alternative forced-choice judgment of the orientation of the second target relative to the first one (clockwise or counterclockwise). Informative feedback was provided after each response by brightening (correct response) or dimming (wrong response) the fixation point briefly, which facilitated learning (Goldhacker, Rosengarth, Plank, & Greenlee, 2014). The next trial began 800 to 1200 ms after feedback. $\Delta\theta$ varied trial by trial and was controlled by QUEST staircases to estimate subjects' discrimination thresholds at 75% correct.

During the three test phases, subjects' orientation discrimination thresholds were measured with four test stimuli: the crowded trained target, the isolated trained target, the crowded untrained target, and the isolated untrained target (Figure 1C, first row). The untrained target was identical to the trained target except that its orientation was perpendicular to that of the trained target. Thirty-two QUEST staircases (same as above), eight for each test stimulus, were completed in a random order. Starting values in the QUEST staircases were identical. During Training 1, subjects continued practicing with the crowded (trained) target until the mean threshold from five consecutive QUEST staircases was lower than the threshold measured with the isolated trained target at Pre. During Training 2, subjects underwent six more daily training sessions with the crowded target.

Experiments 2 and 3 had the same design and trained stimulus as Experiment 1. Two of the four test stimuli (the crowded trained target and the isolated trained target) in Experiment 1 were also used in Experiments 2 and 3. In Experiment 2, the gratings in the crowded untrained target and the isolated untrained target in Experiment 1 were replaced with random-dot kine-matograms (RDKs; radius: 1.5°; dot density: 8/°^{2}; velocity: 10°/s; luminance: 0.01 cd/m^2). The moving direction of the target RDK deviated from the orientation of the trained target in Experiment 1 by 60°, either clockwise or counterclockwise. The directions of two flanker RDKs were randomized (Figure 1C, second row). Similar to the orientation discrimination measurement, we measured subjects' motion direction discrimination thresholds with these two new test stimuli. In Experiment 3, the crowded trained target and the isolated trained target in Experiment 1 were also presented in the upper-right visual quadrant, referred to as the crowded untrained target and the isolated untrained target, respectively (Figure 1C, third row).

Experiment 4 also had the same design as Experiment 1. The trained and test stimuli in Experiment 4 were similar to those in Experiment 1, except that the stimuli were presented at 6° eccentricity and the radius of the target and flanker gratings was reduced to 0.98° according to the cortical magnification factor for matching the cortical representation sizes of the stimuli between Experiments 1 and 4 (Duncan & Boynton, 2003). The center-to-center distance between the target and the flankers was 1.96° (Figure 1C, fourth row).

The design of Experiment 5 was slightly different from that of Experiment 1. It had only three phases: Pre, Training 1, and Mid. The contrast and spatial frequency of the target and flankers were identical to those in Experiment 1, but their radius and center-to-center distance were reduced to half of those in Experiment 1. The stimuli were presented at the same eccentricity as that in Experiment 1 (Figure 1C, fifth row). During Training 1, all subjects underwent 10 daily training sessions with the crowded target.

Data analysis

For each test stimulus, discrimination thresholds from eight QUEST staircases were averaged as a measure of subjects' discrimination performance at Pre, Mid, and Post. Subjects' performance improvements with a test stimulus from Pre to Mid and from Mid to Post were calculated as (*pretraining threshold – midtraining threshold*)/*pre-training threshold* $\times$ *100%*) and (*mid-training threshold – posttraining threshold*)/*midtraining threshold* $\times$ *100%*), respectively. Because the learning effects after training at 67.5° and 157.5° were very similar, they were pooled together for further analysis.

Results

Experiment 1: Perceptual learning with crowded orientation and its transfer to crowded orthogonal orientation

Subjects were trained to perform an orientation discrimination task at 67.5° or 157.5° with a crowded target. Their orientation discrimination thresholds were measured using QUEST staircase throughout training. A daily training session consisted of 30 QUEST staircases of 40 trials. Before training, we measured subjects' orientation discrimination thresholds at Pre with four test stimuli: the crowded trained target, the isolated trained target, the crowded untrained target, and the isolated untrained target (Figure 2A). The orientation of the untrained target was perpendicular to that of the trained target. Subjects' discrimination thresholds were much higher with the crowded targets than with the isolated targets: crowded trained target versus isolated trained target, $t(7) = 10.12$, $p < 0.001$; crowded untrained target versus isolated untrained target, $t(7) = 10.23$, $p < 0.001$, demonstrating that the
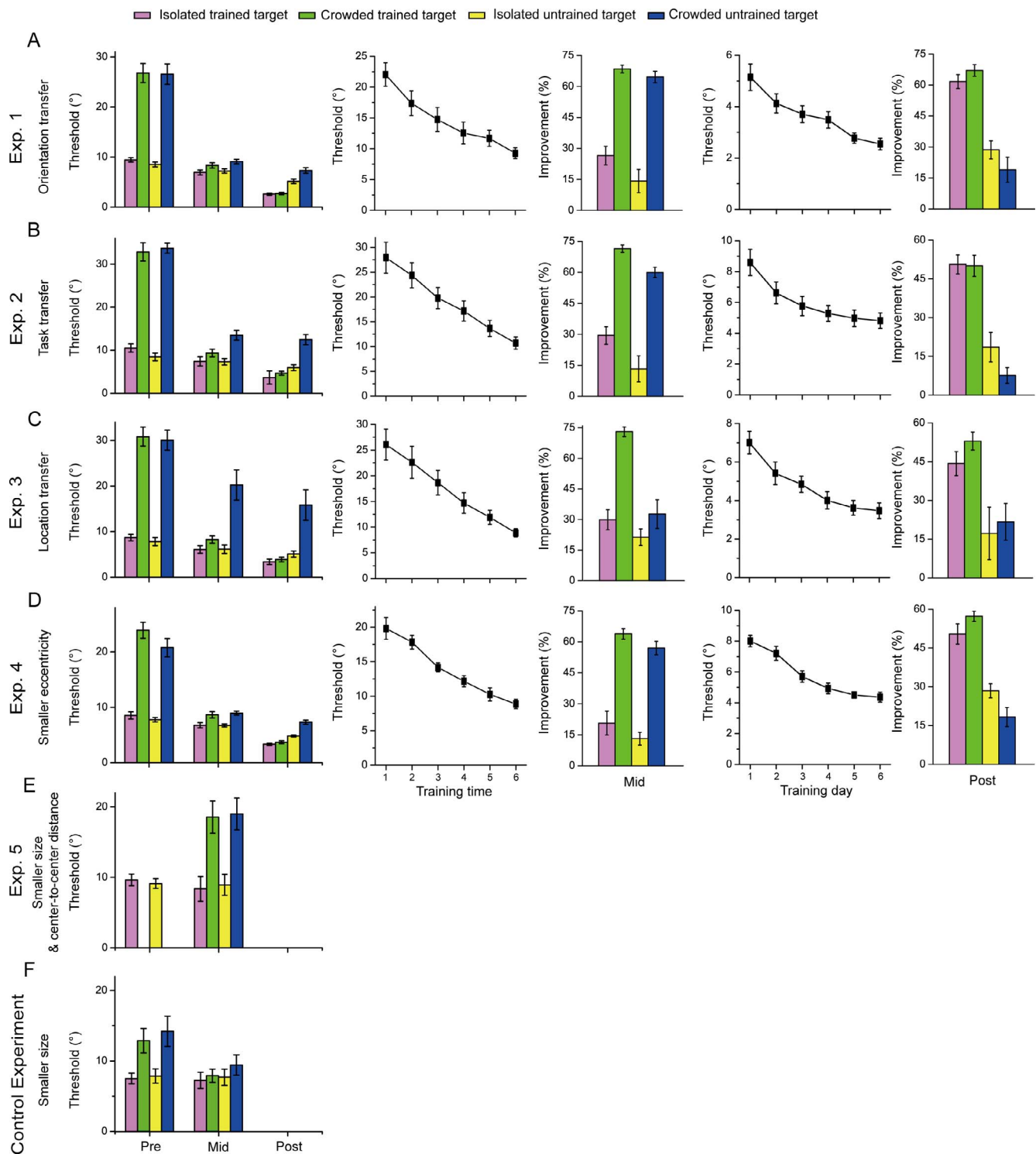
Figure 2. Psychophysical results of Experiments 1–5 and the control experiment. (A–D) First column (from left to right): discrimination thresholds for the four test stimuli at Pre, Mid, and Post. Second column: learning curve during Training 1. For individual subjects, staircases during Training 1 were split into six equally sized bins based on the training progress. The average discrimination threshold in each bin was plotted as a function of bin, referred to as the *learning curve*. Learning curves were then averaged across subjects. Third column: percentage improvements in discrimination performance from Pre to Mid. Fourth column: learning curve during Training 2. Discrimination thresholds are plotted as a function of training day. Fifth column: percentage improvements in discrimination performance from Mid to Post. (E, F) Discrimination thresholds for the four test stimuli at Pre and Mid. Error bars denote 1 SEM across subjects.

presentation of the nearby flankers led to strong crowding.

During Training 1, subjects' performance improved quickly and substantially. The training ceased after $1,760 \pm 302$ trials (about 1.5 training sessions, throughout the article, $X \pm Y$ indicates the mean $\pm$ *SEM* across subjects), because at that time, the mean threshold from the last five QUEST staircases with the crowded trained target was lower than the threshold measured with the isolated trained target at Pre. At Mid, subjects' discrimination thresholds with the four test stimuli were measured again. There was no significant difference between the crowded trained target and the isolated trained target, $t(7) = 2.37$, $p > 0.05$, suggesting that, after Training 1, the crowding effect was completely removed. Then we calculated percentage improvements in discrimination performance from Pre to Mid. The improvements with the crowded trained target ($68.47\% \pm 1.86\%$), the isolated trained target ($26.54\% \pm 4.51\%$), and the crowded untrained target ($64.60\% \pm 2.77\%$) were significant, all $t(7) > 5.61$, $p < 0.001$, but not with the isolated untrained target ($14.22\% \pm 5.64\%$), $t(7) = 2.02$, $p > 0.05$. The difference between the improvements with the isolated trained target and the isolated untrained target was significant, $t(7) = 2.74$, $p < 0.05$. An interesting phenomenon observed here is that the learning effect with the crowded trained target could almost completely transfer to the crowded untrained target, although the orientations of the two targets were orthogonal. However, the transfers to the isolated trained target and the isolated untrained target were weak despite the fact that the isolated trained target owned the trained orientation. In other words, the major effect of Training 1 was the breaking of crowding, rather than sensitivity improvement specific to the trained orientation that was found by many previous perceptual learning studies (Adab & Vogels, 2011; Ghose, Yang, & Maunsell, 2002; Raiguel, Vogels, Mysore, & Orban, 2006; Schoups, Vogels, Qian, & Orban, 2001).

During Training 2, subjects underwent six more daily training sessions with the crowded target. At Post, we measured subjects' discrimination thresholds with the four test stimuli a third time. There was still no significant difference between the crowded trained target and the isolated trained target, $t(7) = 0.81$, $p > 0.05$. The improvements in discrimination performance from Mid to Post were $67.07\% \pm 2.79\%$ for the crowded trained target, $61.68\% \pm 3.36\%$ for the isolated trained target, $19.02\% \pm 6.06\%$ for the crowded untrained target, and $28.64\% \pm 4.36\%$ for the isolated untrained target, all $t(7) > 2.86$, $p < 0.05$. The learning effect with the crowded trained target almost completely transferred to the isolated trained target (these two targets had the trained orientation), whereas

the transfers to the crowded untrained target and the isolated untrained target were weak (the orientation of the two targets was perpendicular to the trained orientation). These results demonstrated that, distinct from Training 1, the effect of Training 2 manifested as improved sensitivity specifically to the trained orientation.

The findings in Experiment 1 showed that perceptual learning with crowded orientation had two distinct stages. In the first stage, subjects learned to break crowding, and the learning effect completely transferred to the orientation orthogonal to the trained orientation, suggesting that subjects might have learned the general ability to separate the target and flankers. This hypothesis was further tested in the following experiments. In the second stage, the learning effect was very similar to many classical perceptual learning effects, exhibiting a hallmark feature of perceptual learning—specificity to the trained feature (i.e., orientation).

Experiment 2: Perceptual learning with crowded orientation and its transfer to crowded motion stimulus

Experiment 2 aimed to examine whether the learned ability to break the orientation crowding could generalize to break motion crowding. The experiment used the same design and stimuli as Experiment 1, except that the targets and flankers in two test stimuli (the crowded untrained target and the isolated untrained target) were RDKs. We measured motion direction discrimination thresholds with the two test stimuli.

At Pre, the crowding effects were very strong for both the orientation stimulus (crowded trained target vs. isolated trained target), $t(8) = 13.50$, $p < 0.001$, and the motion stimulus (crowded untrained target vs. isolated untrained target), $t(8) = 15.77$, $p < 0.001$ (Figure 2B). Similar to the finding in Experiment 1, Training 1 improved subjects' performance quickly and substantially, and it ceased after practicing $1,910 \pm 286$ trials. At this point, the mean threshold from the last five QUEST staircases with the crowded trained target was lower than the threshold measured with the isolated trained target at Pre. At Mid, we measured subjects' orientation or direction discrimination thresholds with the four test stimuli and calculated the percentage improvements in discrimination performance from Pre to Mid. The improvements with the crowded trained target ($71.43\% \pm 1.78\%$), the isolated trained target ($30.73\% \pm 4.26\%$), and the crowded untrained target ($60.07\% \pm 2.49\%$) were significant, all $t(8) > 6.55$, $p < 0.001$, but not with the isolated untrained target ($11.44\% \pm 6.33\%$), $t(8) = 1.78$, $p >$

0.05. The difference between the improvements with the isolated trained target and the isolated untrained target was significant, $t(8) = 3.801$, $p < 0.01$. The learning effect with the crowded trained target could largely transfer to the crowded untrained target, despite that the two stimuli consisted of dramatically different components (i.e., oriented grating and RDK). However, the transfers to the isolated trained target and the isolated untrained target were much weaker.

After Training 2, the improvements in discrimination performance from Mid to Post were 49.04% ± 4.11% for the crowded trained target, 49.27% ± 3.67% for the isolated trained target, 7.62% ± 3.02% for the crowded untrained target, and 16.89% ± 5.68% for the isolated untrained target, all $t(8) > 2.55$, $p < 0.05$. The learning effect with the crowded trained target completely transferred to the isolated trained target. But the transfers to the crowded untrained target and the isolated untrained target were weak.

These findings provided further evidence that, in the first learning stage, subjects learned to separate the target and flankers presented at the trained location. The improved segmentation ability persisted despite the fact that the trained and test stimuli (oriented grating vs. RDK) were completely different. In the second learning stage, the learning effect showed specificity to the trained feature, replicating the finding in Experiment 1.

Experiment 3: Perceptual learning with crowded orientation and its transfer to the opposite visual hemi-field

Experiment 3 was designed to examine whether the learned ability to break crowding could generalize to the opposite visual hemi-field. The experiment used the same design and stimuli as Experiment 1, except that the crowded trained target and the isolated trained target in Experiment 1 were also presented in the upper-right visual quadrant, referred to as the crowded untrained target and the isolated untrained target, respectively.

At Pre, the crowding effects were very strong in both visual hemi-fields, both $t(7) > 12.97$, $p < 0.001$ (Figure 2C). Training 1 ceased after subjects practiced 2,090 ± 407 trials. It improved subjects' performance dramatically and removed the crowding effect in the trained (i.e., left) visual hemi-field. Performance improvements from Pre to Mid were 72.77% ± 2.33% for the crowded trained target, 31.32% ± 4.90% for the isolated trained target, 34.54% ± 7.03% for the crowded untrained target, and 21.18% ± 4.08% for the isolated untrained target, all $t(7) > 4.52$, $p < 0.01$. Different from Experiments 1 and 2, the transfer of the learning effect to the crowded untrained target was weak in Experi-

ment 3, which was comparable to the transfer to the isolated trained target and the isolated untrained target. This finding demonstrated that the improved segmentation ability after Training 1 manifested largely at the trained location.

From Mid to Post, the improvements with the crowded trained target (53.48% ± 3.48%), the isolated trained target (44.87% ± 4.66%), and the crowded untrained target (22.78% ± 7.12%) were significant, all $t(7) > 2.92$, $p < 0.05$, but not with the isolated untrained target (12.85% ± 10.16%), $t(7) = 1.52$, $p > 0.05$. Again, this finding demonstrated that the learning effect from Training 2 exhibited specificity for the trained orientation at the trained location.

Experiment 4: Perceptual learning with crowded orientation at smaller eccentricity

Experiment 4 examined whether the results in Experiment 1 could be replicated at 6° eccentricity. The stimuli in Experiment 1 were reduced in size according to the cortical magnification factor and then used in Experiment 4. At Pre, the crowding effects were very strong, both $t(7) > 7.11$, $p < 0.001$ (Figure 2D). Training 1 ceased after subjects practiced 1,720 ± 418 trials. From Pre to Mid, the improvements in discrimination performance were 63.39% ± 2.56% for the crowded trained target, 19.01% ± 5.76% for the isolated trained target, 55.43% ± 3.28% for the crowded untrained target, and 12.55% ± 3.10% for the isolated untrained target, all $t(7) > 3.04$, $p < 0.05$. From Mid to Post, the improvements were 57.20% ± 1.95% for the crowded trained target, 49.14% ± 3.94% for the isolated trained target, 18.00% ± 3.71% for the crowded untrained target, and 28.02% ± 2.72% for the isolated untrained target, all $t(7) > 4.22$, $p < 0.01$. The two-stage learning effects were very similar to those in Experiment 1.

Experiment 5: Limited effect of perceptual learning with crowded orientation

In Experiment 5, the stimulus sizes were reduced to half of those in Experiment 1. The stimuli were still presented at the same eccentricity as that in Experiment 1. We examined whether crowding could be completely eliminated with smaller stimuli. At Pre, crowding effects were too strong to measure subjects' orientation discrimination thresholds with the crowded target (not reported in Figure 2E). Subjects' responses to a 90° orientation difference between two crowded targets were at chance level, although isolated targets could be well discriminated.

During Training 1, subjects learned to perform the discrimination task with the crowded target. However, even after 10 days' training, the crowding effect could not be completely eliminated, and training ceased. At Mid, the orientation discrimination thresholds with the isolated target were not significantly different from those at Pre, both $t(9) < 0.73$, $p > 0.05$. Although training decreased the orientation discrimination thresholds with the crowded targets, the thresholds were still much higher than those with the isolated targets, both $t(9) > 4.45$, $p < 0.01$. Similar to Experiment 1, the threshold with the crowded trained target was not significantly different from that with the crowded untrained target, $t(9) = 0.33$, $p > 0.05$. These results suggested that, when the target and flankers were too close, training could only reduce, but not eliminate, crowding.

Relative to Experiment 1, both the size of the target and flankers and their center-to-center distance were reduced in Experiment 5. Both factors might be able to explain the results of Experiment 5. To investigate which factor caused the lack of the training effect in Experiment 5, we added a control experiment performed with four subjects. In the control experiment, the grating size was the same as that in Experiment 5. The center-to-center distance of the gratings and the eccentricity of the target were identical to those in Experiment 1. Thus, the flankers and target were no longer abutting. At Pre, the crowding effects were significant, both $t(3) > 3.77$, $p < 0.05$. Training 1 ceased after subjects practiced $2,100 \pm 439$ trials. From Pre to Mid, the improvements in discrimination performance were $38.78\% \pm 2.30\%$ for the crowded trained target, $3.52\% \pm 2.90\%$ for the isolated trained target, $33.87\% \pm 7.91\%$ for the crowded untrained target, and $2.08\% \pm 4.13\%$ for the isolated trained target. At Mid, there was no significant difference between the crowded trained target and the isolated trained target, $t(4) = 1.61$, $p > 0.05$, demonstrating that after Training 1, the crowding effect was completely removed. This finding suggests that the center-to-center distance between the target and flankers plays a major role in the breaking of crowding.

In Experiments 1 to 5, there might be some retest effects due to practice (i.e., threshold measurement) at Pre. We recruited two new subjects to measure the retest effects. The test-retest experiment was identical to Experiment 1 except that there was no intervening training. We measured orientation discrimination thresholds twice, with a 3-day gap between two measurements. Four stimuli were retested, including the isolated 67.5° target, the crowded 67.5° target, the isolated 157.5° target, and the crowded 157.5° target. Retest improvements for the four stimuli were 17.29%, 23.21%, 13.98%, and 30.02%, respectively. The retest effects were much smaller than the learning effects reported in Experiment 1.

Discussion

We performed a series of psychophysical experiments to investigate perceptual learning effects with crowded stimuli. Subjects were trained to perform an orientation discrimination task with a crowded target. Perceptual learning with a crowded orientation exhibited two distinct temporal stages. In the first (early) stage, which was relatively fast, learning broke crowding and the learning effect could transfer to a different crowded orientation, even a different type of stimulus (i.e., crowded motion). However, the transfer to the opposite visual hemi-field was weak. In the second (late) stage, subjects' performance gradually improved and showed specificity to the trained orientation. These findings shed new light on the mechanisms of visual crowding and perceptual learning as discussed below.

Natural visual scenes are usually cluttered. In such scenes, many objects in the periphery are crowded and difficult to identify, simply because of the dense array of clutter. It is well known that training can improve performance for many visual tasks (Sagi, 2011; Watanabe & Sasaki, 2015). Consistent with this, several previous studies (Chung, 2007; Huckauf & Nazir, 2007; Hussain et al., 2012; Sun et al., 2010; Xiong et al., 2015) trained human subjects on a crowded letter identification task and found that crowding could be ameliorated by training. Performance improvements ranged from 30% to 88%, depending on the amount of training. However, in all these previous studies, the crowding effect could only be alleviated but not completely eliminated. Our current study went beyond this finding by showing that, in some situations, crowding can be completely eliminated by training. Furthermore, the transfer properties of breaking crowding inform us about the underlying neural mechanisms of crowding. Different stimuli and tasks between our study and the previous ones could explain the discrepancy. Especially, the size of the target and flankers and their spacing might be important factors. In the previous studies, both of them were much smaller than those in the current study, which might prevent crowding effect being completely eliminated, as suggested by Experiment 5.

Crowding is often attributed to inappropriate integration (or pooling) of signals over space because peripheral vision lacks sufficient spatial resolution to discern a target and its flankers (Levi et al., 2002; Pelli & Tillman, 2008). When faced with crowded stimuli, a strategy for the visual system to quickly improve

performance with a crowded target is to learn to ignore or suppress the information from flankers that are irrelevant, and may even be distracting, to the task of identifying or discriminating the target. To do so, the visual system needs to acquire the ability to segment the target and flankers and then individuate and access the target. Indeed, in the early learning stage, subjects quickly learned to break crowding. Moreover, the generalization of breaking crowding to the perpendicular orientation and the motion stimulus provided key evidence for this segmentation idea. Because the learning effect uncovered in the early stage is independent of stimulus type, it is likely that what subjects had learned is isolating and accessing the area occupied by the target. How does the brain implement this? One possibility is that the brain learns to improve the resolution of spatial attention. It has been proposed that crowding could be ascribed to coarse resolution of spatial attention (He et al., 1996) or unfocused spatial attention (Strasburger, 2005). When the target and flankers are spaced more finely than the limit of attentional resolution, the target cannot be selected individually for further processing, resulting in crowding. In terms of the attention resolution theory, our finding here can be simply explained as a result of our subjects being more capable of focusing their attention toward the target instead of dispersing their attention over the flankers. Once subjects' attentional spotlight was shrunk by training to a certain size to just cover the target area, interference from the flankers could be suppressed or ignored, leading to the breaking of crowding. A related explanation of the breaking of crowding is that training locally inhibits activity at the flanker locations, reducing the interference from the flankers consequently.

Although the training-induced change of attentional resolution provides a plausible explanation for the transfer of breaking crowding to the perpendicular orientation and the motion stimulus, a seemingly paradoxical finding here is that the transfer of breaking crowding to the opposite visual hemi-field was weak. Traditionally, attention is thought of as a centrally organized process that controls selection similarly along the entire information-processing stream in the brain (Broadbent, 1958; Moran & Desimone, 1985). Thus, we expected to find a complete transfer between the left and right visual hemi-fields. Recent psychophysical and electroencephalography studies, however, demonstrated that attentional mechanisms were fundamentally constrained by anatomical properties of visual cortical areas. For example, it was easier to track multiple targets across the left and right visual hemi-fields than within the same visual hemi-field (Alvarez, Gill, & Cavanagh, 2012; Chakravarthi & Cavanagh, 2009). The benefit of dividing attention across separate visual hemi-fields emerged at an early sensory level

(Störmer, Alvarez, & Cavanagh, 2014). Similarly, Carlson, Alvarez, and Cavanagh (2007) found that tracking performance improved when target objects appeared in separate visual quadrants compared with when they appeared the same distance apart but within a single quadrant. Consistent with these studies, our findings here suggest that the trained-induced change of attentional resolution might reflect plasticity of the higher-level attention network (Corbetta & Shulman, 2002), which was further constrained by anatomical properties of lower-level cortical areas.

Recently, Sun et al. (2010) used ideal observer analysis and a training paradigm to identify the functional mechanism of crowding. They suggest that the mechanism underlying the reduction of crowding following training is attributable to the perceptual window being more capable of adjusting its size to gather relevant information from the target. After training, subjects with inappropriately large windows reduced their window size to exclude interference from flankers. The window size can be quantified as the critical distance of crowding (Bouma, 1970). In the current study, because learning to break crowding was quick, there were not enough trials for measuring the critical distance. The notion of the perceptual window is also consistent with what Pelli et al. (2007; Pelli & Tillman, 2008) referred to as "isolation field" or "combination field." Although these early ideas are generally in accordance with our explanation, however, without having performed the transfer tests here, it would be difficult to speculate the cortical mechanisms underlying the reduction of crowding.

It should be noted that the breaking of crowding occurred only when the target and flankers were separated beyond a certain distance. As demonstrated by Experiment 5, when the target and flankers were too close, although the crowding effect could be reduced by training, it could not be completely removed. According to the attention resolution theory, this is because, even after intensive training, the attention resolution was still not fine enough to select the target individually for further processing based on its location, and the interference from flankers could not be suppressed or ignored. Crowding is a form of inhibitory interaction. Recent brain imaging studies (J. Chen et al., 2014; Kwon, Bao, Millin, & Tjan, 2014; Millin, Arman, Chung, & Tjan, 2013) demonstrated that crowding manifested as an attention-dependent suppressive cortical interaction between the target and flankers in early visual areas. Based on the findings in the current study, we speculate that, if a finer attention resolution following training could grab onto the target and enhance the processing of the target, the suppressive interaction could be counteracted, and thereby crowding could be removed. On the other hand, if the attention resolution is still too coarse, the suppressive

interaction depending on the distance between the target and flankers may play a major role in determining the magnitude of crowding. Taken together, crowding is determined by the combination of constraints at multiple levels of cortical processing, including low-level cortical interaction and high-level attention.

The performance improvement in the early training stage was largely due to the improved general ability of segmenting the target and flankers, which manifested with the crowding configuration (i.e., the radial configuration) used in the study. Distinct from the early training stage, the improvement in the late training stage was mainly attributed to the perceptual learning effect specific to the trained orientation. The visual system might have learned to refine the neural representation of the trained orientation in sensory areas and/or improve relevant decision-making processes in higher cortical areas (N. Chen et al., 2015; Law & Gold, 2008; Schoups et al., 2001). It is noteworthy that, in the early training stage, there was a difference between the improvements with the isolated trained target and the isolated untrained target, suggesting that some orientation-specific learning might have occurred. Therefore, these two training stages are not mutually exclusive. This two-stage learning process illustrates a learning strategy for our brain to deal with the notoriously difficult problem of recognizing peripheral objects in cluttered visual scenes. The brain chooses to solve the "easy and general" part of the problem first, then tackle the "difficult and specific" part afterward. This process is in accordance with the reverse hierarchy theory of perceptual learning (Ahissar & Hochstein, 1997, 2004), which claims that learning proceeds as a countercurrent along the cortical hierarchy, with high-level easy-condition learning occurring before low-level hard-condition learning. Our findings are also consistent with previous works showing increased specificity of learning with more practice (Hung & Seitz, 2014; Jeter, Dosher, Liu, & Lu, 2010).

In sum, we took advantage of the perceptual learning paradigm to investigate the mechanisms of visual crowding and revealed a previously unknown two-stage learning process to break crowding. Given that crowding can be reduced and even completely eliminated by a relatively short period of training, training effects should be taken into consideration when researchers study the mechanisms of crowding. In the future, the breaking of crowding should be investigated with various brain imaging and neurophysiological techniques to fully uncover its underlying neural mechanisms, which will contribute significantly to our understanding of object recognition, scene analysis, and even conscious awareness.

Acknowledgments

References

Adab, H. Z., & Vogels, R. (2011). Practicing coarse orientation discrimination improves orientation signals in macaque cortical area v4. *Current Biology*, *21*, 1661–1666.

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

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

Alvarez, G. A., Gill, J., & Cavanagh, P. (2012). Anatomical constraints on attention: Hemifield independence is a signature of multifocal spatial selection. *Journal of Vision*, *12*(5):9, 1–20, doi:10.1167/12.5.9. [PubMed] [Article]

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.

Broadbent, D. E. (1958). *The effects of noise on behaviour*. Elmsford, NY: Pergamon Press.

Butler, T. W., & Westheimer, G. (1978). Interference with stereoscopic acuity: Spatial, temporal, and disparity tuning. *Vision Research*, *18*, 1387–1392.

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences, USA*, *104*, 13496–13500.

Chakravarthi, R., & Cavanagh, P. (2009). Bilateral field advantage in visual crowding. *Vision Research*, *49*, 1638–1646.

Chung, S. T. (2007). Learning to identify crowded

letters: does it improve reading speed? *Vision Research*, *47*, 3150–3159.

Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., & Fang, F. (2014). Attention-dependent early cortical suppression contributes to crowding. *Journal of Neuroscience*, *34*, 10465–10474.

Chen, N., Bi, T., Zhou, T., Li, S., Liu, Z., & Fang, F. (2015). Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *Neuroimage*, *115*, 17–29.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review Neuroscience*, *3*, 201–215.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*, 659–671.

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *Journal of Neurophysiology*, *87*, 1867–1888.

Goldhacker, M., Rosengarth, K., Plank, T., & Greenlee, M. W. (2014). The effect of feedback on performance and brain activation during perceptual learning. *Vision Research*, *99*, 99–110.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences*, *1*, 86–93.

Huckauf, A., & Nazir, T. A. (2007). How odgcrnwi becomes crowding: Stimulus-specific learning reduces crowding. *Journal of Vision*, *7*(2):18, 1–12, doi:10.1167/7.2.18. [PubMed] [Article]

Hung, S., & Seitz, A. R. (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *Journal of Neuroscience*, *34*, 8423–8431.

Hussain, Z., Webb, B. S., Astle, A. T., & McGraw, P. V. (2012). Perceptual learning reduces crowding in amblyopia and in the normal periphery. *Journal of Neuroscience*, *32*, 474–480.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216.

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

Kwon, M. Y., Bao, P., Millin, R., & Tjan, B. S. (2014). Radial-tangential anisotropy of crowding in the early visual areas. *Journal of Neurophysiology*, *112*, 2413–2422.

Law, C. T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*, 505–513.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*, 635–654.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, *2*(2):3, 167–177, doi:10.1167/2.2.3. [PubMed] [Article]

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*, 963–977.

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, *7*(2):24, 1–11, doi:10.1167/7.2.24. [PubMed] [Article]

Millin, R., Arman, A. C., Chung, S. T. L., & Tjan, B. S. (2013). Visual crowding in V1. *Cerebral Cortex*, *7*, 1–9.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*, 1129–1135.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]

Raiguel, S., Vogels, R., Mysore, S. G., & Orban, G. A. (2006). Learning to see the difference specifically alters the most informative V4 neurons. *Journal of Neuroscience*, *26*, 6589–6602.

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

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*, 549–553.

Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11):8, 1024–1037, doi:10.1167/5.11.8. [PubMed] [Article]

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*, 495–508.

Störmer, V. S., Alvarez, G. A., & Cavanagh, P. (2014). Within-hemifiled competition in early visual areas limits the ability to track multiple objects with attention. *Journal of Neuroscience*, *34*, 11526–11533.

Sun, G. J., Chung, S. T., & Tjan, B. S. (2010). Ideal observer analysis of crowding and the reduction of crowding through learning. *Journal of Vision*, *10*(5): 16, 1–14, doi:10.1167/10.5.16. [PubMed] [Article]

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology*, *1*, 197–221.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.

Westheimer, G., Shimamura, K., & McKee, S. P. (1976). Interference with line-orientation sensitivity. *Journal of the Optical Society of America*, *66*, 332–338.

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.

Xiong, Y., Yu, C., & Zhang, J. (2015). Perceptual learning eases crowding by reducing recognition errors but not position errors. *Journal of Vision*, 15(11):16, 1-13, doi:10.1167/15.11.16. [PubMed] [Article]