

# Perceptual learning of ensemble and outlier perception

Shaul Hochstein

ELSC Safra Brain Research Center and Life Sciences  
Institute, Hebrew University, Jerusalem, Israel



Marina Pavlovskaya

Lowenstein Rehabilitation Hospital and Tel Aviv  
University, Tel Aviv, Israel



**Previous studies have demonstrated a complex relationship between ensemble perception and outlier detection. We presented two array of heterogeneously oriented stimulus bars and different mean orientations and/or a bar with an outlier orientation, asking participants to discriminate the mean orientations or detect the outlier. Perceptual learning was found in every case, with improved performance accuracy and speeded responses. Testing for improved accuracy through cross-task transfer, we found considerable transfer from training outlier detection to mean discrimination performance, and none in the opposite direction. Implicit learning in terms of increased accuracy was not found in either direction when participants performed one task, and the second task's stimulus features were present. Reaction time improvement was found to transfer in all cases. This study adds to the already broad knowledge concerning perceptual learning and cross-task transfer of training effects.**

At the same time, there is considerable debate as to learning specificity versus transfer of learning effects to new conditions of the same task, including limited transfer between locations (Ahissar & Hochstein, 1993; Karni & Sagi, 1991), across hemispheres (Pavlovskaya & Hochstein, 2011), between eyes (Ball & Sekuler, 1987; Karni & Sagi, 1991; Karni & Sagi, 1993), between spatial frequencies (Fiorentini & Berardi, 1981), between orientations (Fiorentini & Berardi, 1981), and across tasks (Ahissar & Hochstein, 1993; Treisman, Vieira, & Hayes, 1992). For reviews, see Goldstone, 1998; Holtmaat and Svoboda, 2009; Lu, Yu, Sagi, Watanabe, and Levi, 2009; Sagi, 2011; Sale, Berardi, and Maffei, 2009; Sasaki, Nanez, and Watanabe, 2010; and chapters in Fahle and Poggio, 2002, and in this special issue.

The question of specificity versus transfer has become especially poignant due to two issues: reverse hierarchy theory suggested that the cerebral site of initial conscious perception (Hochstein & Ahissar, 2002) and of initial perceptual learning (Ahissar & Hochstein, 1997; Ahissar & Hochstein, 2004; see also Ahissar et al., 2008, Treisman, 2009) are both at high cortical levels, and these guide further recruitment of later, lower level involvement, including the so-called “eureka” effect. This theory thus upends discussion of what is early and what is late. Second, there is recent debate concerning perceptual learning without attention to the task (Sasaki Nanez & Watanabe, 2010; Seitz & Watanabe, 2003; Tsushima, Seitz, & Watanabe, 2008; Watanabe, Nanez, & Sasaki, 2001).

Another perceptual phenomenon has received considerable interest since the beginning of this century, namely, ensemble perception. It turns out that observers are very good at discerning summary statistics of sets of stimulus elements, including average parameters of a group of elements, and the range (or variance) of these parameters. Such summary statistics are rapidly extracted from sets of similar items, presented spatially (Alvarez & Oliva, 2009; Ariely, 2001) or temporally (Corbett & Oriet, 2011; Gorea, Belkoura, & Solomon, 2014; Hubert-Wallander & Boynton, 2015).

## Introduction

Much has been said about perceptual learning of performance of basic-level perceptual tasks. There seems to be ubiquitous learning due to practice, including for outlier detection (classically called feature search or “pop out” by Treisman & Gelade, 1980; see Ahissar & Hochstein, 1993; Ahissar & Hochstein, 1996; Ahissar & Hochstein, 1997; Ahissar & Hochstein, 2000; Ahissar & Hochstein, 2004; Ahissar, Laiwand, & Hochstein, 2001; Zhaoping, 2009), motion detection (Ball & Sekuler, 1982; Ball & Sekuler, 1987; Dick, Ullman, & Sagi, 1987; Lu, Qian, & Liu, 2004; Thompson & Liu, 2006), texture discrimination (Braun & Sagi, 1991; Husk, Bennett, & Sekuler, 2007; Karni & Sagi, 1991; Karni & Sagi, 1993; Ofen, Moran, & Sagi, 2007; Rubenstein & Sagi, 1990), face identification (Husk, Bennett, & Sekuler, 2007), and other perceptual processes (see articles in this special issue).

Citation: Hochstein, S., & Pavlovskaya, M. (2020). Perceptual learning of ensemble and outlier perception. *Journal of Vision*, 20(8):13, 1–17, <https://doi.org/10.1167/jov.20.8.13>.



Ensemble perception has been studied for basic parameters, including size (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014; Ariely, 2001; Corbett & Oriet, 2011; Morgan, Chubb, & Solomon, 2008; Solomon, 2010), orientation (Alvarez & Oliva, 2009; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018), brightness (Bauer, 2009), spatial position (Alvarez & Oliva, 2008), and speed and direction of motion (Sweeny, Haroz, & Whitney, 2013). Summary statistics perception appears to be a general mechanism operating on various stimulus attributes, including these low-level parameters as noted above, and complex characteristics, such as facial expression or emotion and gender (Haberman & Whitney, 2007; Haberman & Whitney, 2009; Neumann, Schweinberger, & Burton, 2013), object lifelikeness (Yamanashi Leib, Kosovicheva, & Whitney, 2016), biological motion of human crowds (Sweeny, Haroz, & Whitney, 2013), numerical averaging (Brezis, Bronfman, & Usher, 2015) and even category membership (Khayat & Hochstein, 2019; Hochstein, Khayat, Pavlovskaya, Bonneh, Soroker, & Fusi, 2019) for recent reviews, see Bauer, 2015; Cohen et al., 2016; Haberman & Whitney, 2012; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015; Whitney & Yamanashi Leib, 2018; and an upcoming *Attention, Perception, and Psychophysics* special issue).

It was found that that extraction of set summary statistics is automatic, on-the-fly trial-by-trial, and implicit when observers are performing another task, to which they presumably turn their attention (Khayat & Hochstein, 2018; Khayat & Hochstein, 2019). Examples of the methods used in previous studies are shown in Figure 1; see methodologic details in the figure caption.

We now ask if ensemble perception too is malleable and subject to improvement by perceptual learning. Because ensemble perception is so widespread, finding perceptual learning here would expand the range of this learning phenomenon. Furthermore, because ensemble perception can be implicit, when attention is directed to another task, training-based improvement here would bear on the issue of perceptual learning without task performance.

We use the methodology of Hochstein et al. (2018), testing discrimination of the mean orientations of two arrays of short black bars with heterogeneous orientations (on a gray background), or detection of a bar with an outlier orientation within one of these arrays. In each trial, two arrays are presented, sequentially, each containing 69 stimulus bars with a range of orientations. In the mean discrimination task, observers reported which array had the more clockwise mean orientation, and in the outlier detection task they reported which array had a bar with an orientation that was beyond the range of orientations of that array. Hochstein et al. (2018) found that observers were very good at mean orientation discrimination when the two arrays differed sufficiently in mean orientation,

irrespective of the variance of orientations within each array (if the variance was not too great). Detection of the outlier was easy when the outlier orientation differed sufficiently from the edge of the range of array orientations, again irrespective of the variance of the orientations within the arrays.

We now ask if repeated performance of these tasks leads to improved accuracy and/or reduced reaction time (RT). Furthermore, using very similar stimuli for the two tasks, will performance of one task lead to improved performance of the other, perhaps owing to acquaintance with the stimulus arrays or other task paradigm details? Finally, we present array pairs with both different mean orientations, and with one containing an outlier, directing observers to perform one task or the other (i.e., mean discrimination or outlier detection), and ask if performance of one task, with implicit presence of the other, will affect subsequent performance of the nonperformed task.

## Methods

### Participants

Experiments were performed by five in-house volunteer observers, students, and coworkers at the Lowenstein Rehabilitation Center (age 23–28 years; 2 men, 3 women). All were naive as to the goals of the study and had normal or corrected-to-normal vision. We also tested 24 observers on the Amazon Mechanical Turks (MTurk) platform. We selected MTurk observers who would participate in a series of experiments and followed each one's performance over the series of sessions. Participants in all experiments performed one session per day. We have less control of the identities and characteristics of these observers and their precise experimental conditions. Still, we found similar results for these observers and for our in-house laboratory-performed experiments, so that the MTurk results confirm the latter results with a larger group of observers and under a variety of experimental paradigms. We believe that there is benefit in combining in-house and MTurk participants. Following an initial session, we rejected other MTurk participants when there were indications that they were not performing the task, when accuracy rates were at 50% chance level even for easy conditions, and/or RTs were inappropriate (<100 ms or >3 s).

The study was approved by the ethics (Helsinki) committee at the Lowenstein Rehabilitation Center, Raanana, Israel, and participants gave informed consent to participate.

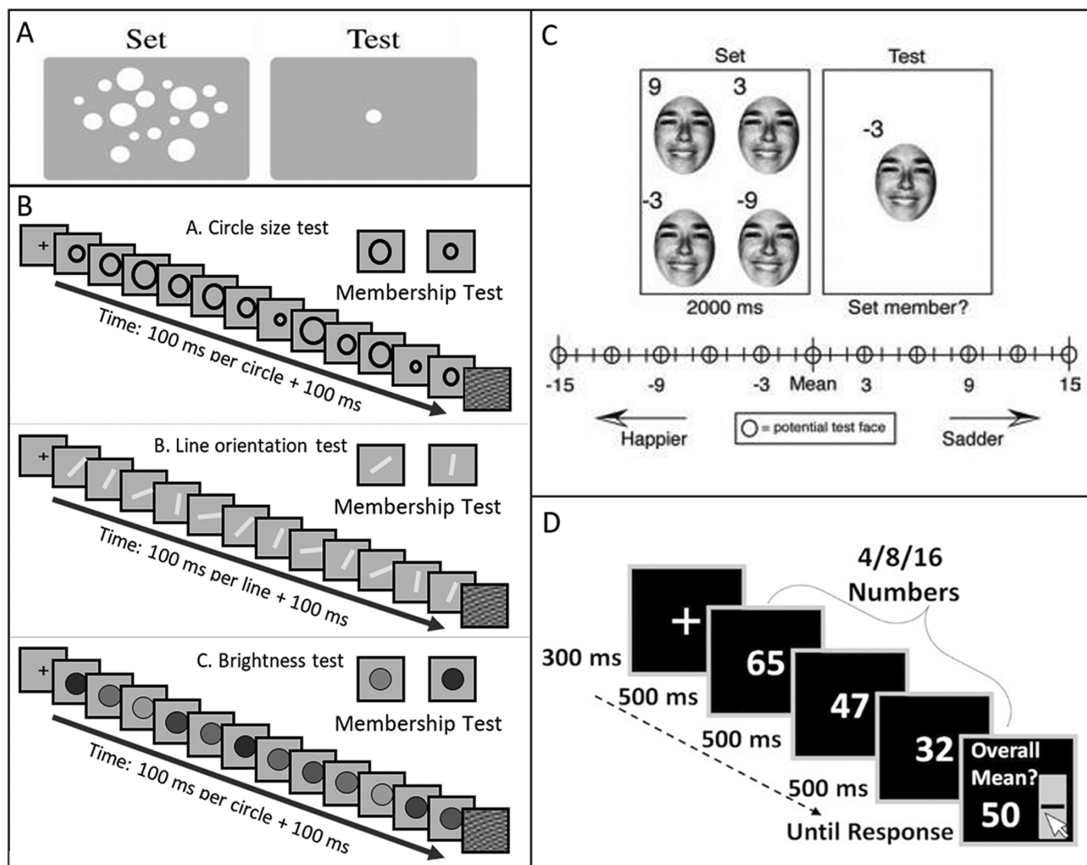


Figure 1. Previous study stimulus sets. (A) Ariely's (2001) representation of the two intervals of his experimental trials. Observers were exposed for 500 ms to a set of spatially dispersed circles differing in size, and then asked if a test stimulus size had been present in the set, or is smaller/larger than the set mean. (B) Khayat & Hochstein's (2018) RSVP sequences of 12 elements, 100 ms each, 100 ms interstimulus interval, followed by a two-alternative membership test, that is, which test element was present in the sequence. Blocks contained circles differing in size, lines differing in orientation, or discs differing in brightness. Observers were unaware that either test element could equal the set mean or the nonmember could be outside the set range, contingencies that were found to affect responses. (C) Haberman & Whitney's (2009) task included four faces (from a set of 4, 8, 12 or 16), differing in facial emotional expression, 2 second presentation. Observers indicated whether the test face was a set member or was happier or sadder than the set mean. (D) Brezis, Bronfman & Usher's (2015) trials consisted of 4, 8, or 16 two-digit numbers sequentially presented at 500 ms/stimulus. Participants estimated set average.

## Experimental set-up

Stimuli were displayed on a 19-in. CRT monitor controlled by dedicated OpenGL-based (Austin, TX) software running on a Windows PC. Video format was true color RGB, 100-Hz refresh rate, with 1024 x 768 pixel resolution. Luminance values were gamma corrected, and mean luminance was approximately 30 cd/m<sup>2</sup>. Sitting distance was 0.7 m, and experiments were administered in near darkness. MTurk programs were written using Adobe Flash (Adobe, Inc., San Jose, CA). As mentioned, we have less control of the precise MTurk participant experimental conditions, including their computer monitors, room

lighting, and sitting distance. The similarity of the results confirms their robustness.

Participants were asked to fixate a central circle, and brief displays prevented scanning eye movements. One reason for conducting the experiment in successive mode was to allow observers to fixate the array center during both array presentations.

## Stimuli and procedure

After the observer initiated the trial by pressing the central mouse key (down arrow for MTurk), a single fixation circle appeared at the center of the monitor,

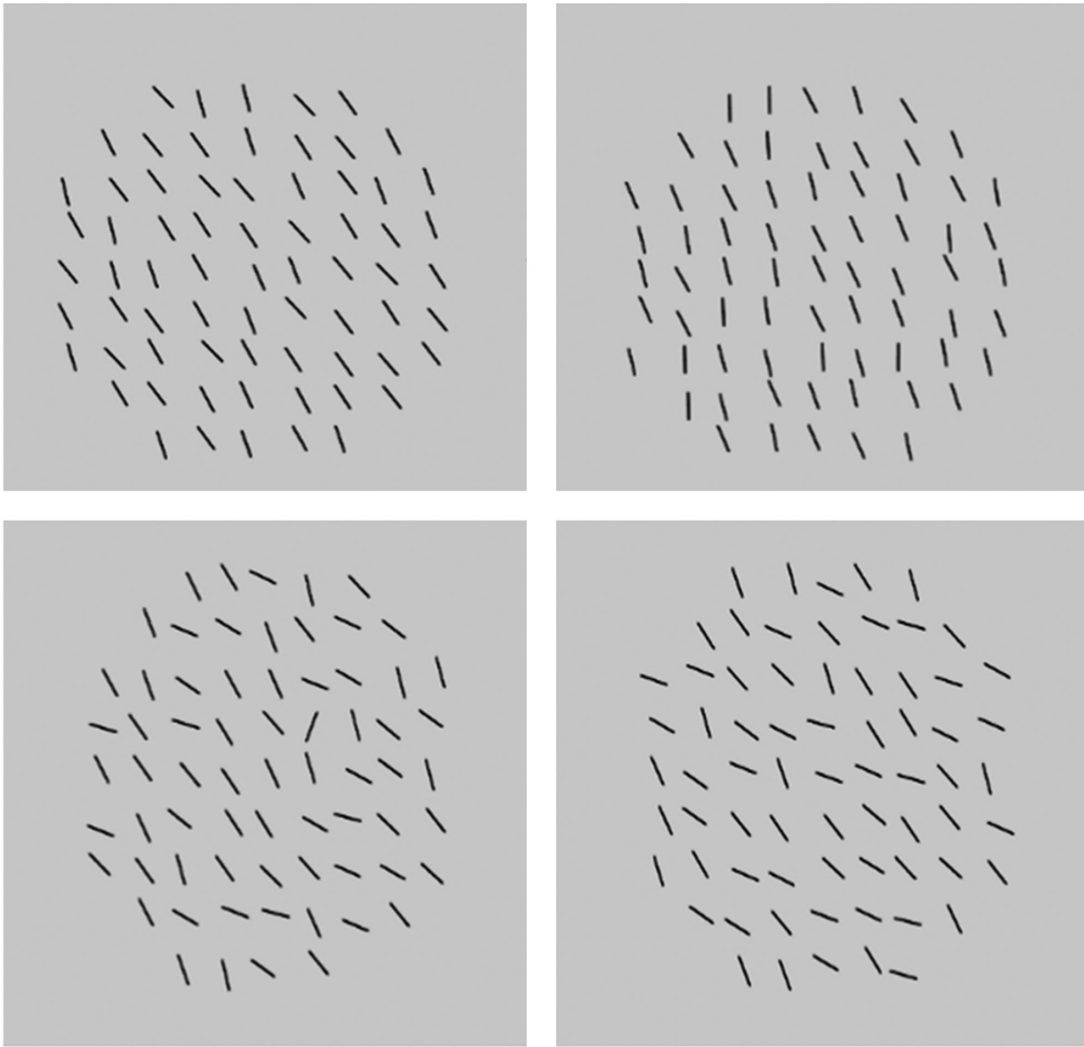


Figure 2. Participants viewed two circular arrays of bars of heterogeneous orientations, presented sequentially in the center of the computer monitor. Top, right, and left: mean orientation discrimination task; observers reported which set mean is tilted more clockwise (in the example, the right one). Bottom, right and left: outlier detection task; observers reported which set contains an outlier (the left one).

with a diameter of  $1.1^\circ$ . After a second keypress, the fixation circle disappeared and the two test arrays were presented in the center of the screen, successively (150 ms each presentation; 300 ms interstimulus interval), unmasked, as demonstrated schematically in the left or right, top or bottom of [Figure 2](#). Arrays were  $6^\circ$  in diameter. Each array contained 69 dark bars arranged in a  $9 \times 9$  grid excluding three in each corner. Bar positions were jittered by up to  $0.2^\circ$  to avoid array homogeneity. Bar length  $\times$  width was  $0.7^\circ \times 0.05^\circ$  and bar orientation was  $60^\circ$  or  $70^\circ$  mean  $\pm$  a random fraction of the variation factor (VAR), which served as the first variable of the study ( $0^\circ$  is horizontal;  $90^\circ$  is vertical). The use of two randomly interleaved mean bar orientations assured that observers were unable to depend on a learned anchor orientation for their judgments ([Ahissar, Lubin,](#)

[Putter-Katz, & Banai, 2006](#)). Note that VAR is exactly the set half-range.

The VAR was set to be  $4^\circ$ ,  $8^\circ$ ,  $16^\circ$  or  $32^\circ$ . For VAR =  $32^\circ$ , there are 65 orientations (with  $1^\circ$  steps) in the range  $(60^\circ - \text{VAR})$  to  $(60^\circ + \text{VAR})$ , that is,  $28^\circ$  to  $92^\circ$  or  $(70^\circ - \text{VAR})$  to  $(70^\circ + \text{VAR})$ ,  $38^\circ$  to  $102^\circ$ , and these were placed randomly in 65 of the 69 bar positions. For lower values of VAR, we used randomly placed multiple repetitions of these values. The final four positions had randomly chosen orientations, chosen as two pairs, equally greater and less than  $60^\circ$  (or  $70^\circ$ ), in order not to change the mean. This random placement was done independently for the two arrays of the display. We only used VAR =  $32^\circ$  for initial testing and do not report these results here. As found by [Hochstein et al. \(2018\)](#), performance depends little on VAR; we therefore average performance data over VAR.

In the orientation discrimination test experiment, all the orientations of one of the arrays (and therefore their mean) were rotated by a variable amount, ORDA (ORientation Difference, Arrays), the second experimental variable. Experimental sessions included trials with ORDA set to be  $\pm 2^\circ$ ,  $\pm 4^\circ$ ,  $\pm 6^\circ$ ,  $\pm 8^\circ$ , or  $\pm 12^\circ$ , in random interleaved order. Observers were instructed to respond by clicking the left or right mouse button when perceiving the first or second array as having a more clockwise rotation, respectively.

In the outlier detection test experiment, the two arrays had the same mean orientation (again, randomly  $60^\circ$  or  $70^\circ$ ), but one of the arrays had an outlier bar, which had an orientation that differed from the mean of the arrays by the variable ORDO (ORientation Difference, Outlier). The outlier could appear in any location within the array, excluding the outer rim and the central position or central  $5 \times 5$  positions, choosing 12 locations to be tested in each session. Experimental sessions included trials with ORDO set to be  $\pm 15^\circ$ ,  $\pm 20^\circ$ , or  $\pm 30^\circ$ , in random interleaved order (ORDO =  $15^\circ$  with VAR =  $16^\circ$  is the only case in which the “outlier” is within the range of the array orientations, so not really an outlier). Observers clicked the left or right mouse button when detecting the outlier in the first or second array, respectively.

In-house observers participated in eight sessions (two observers) or 10 sessions (three observers), four or five for mean orientation discrimination and four or five for outlier detection, performing both tasks in a single sitting with a coffee break between them, or the two tasks in successive days. Each session of orientation discrimination included 450 or 600 trials, three VARs, five ORDAs, 30 or 40 trials/data point/participant. Sessions of outlier detection included 432 or 576 trials, three VARs, three ORDOs, 48 or 64 trials/data point/participant, including 12 or 24 outlier positions.

The MTurk groups performed both tasks in a single sitting, with each session including orientation discrimination: 450 trials, three VARs, five ORDAs, 30 trials/data point/participant; outlier detection: 432 trials, three VARs, three ORDOs, 48 trials/data point including 12 outlier positions.

## Results

### Experiment 1. Ensemble perception learning: Mean discrimination and outlier detection

We tested ensemble perception of 21 observers performing mean discrimination or outlier detection. Thirteen observers (seven men, six women; five in-house and eight MTurk) performed the two tasks every day or two, and we measured performance improvement

over up to eight sessions (four for three in-house observers). An additional four observers performed only mean discrimination, and another four observers performed only outlier detection (five sessions each; three men; five women). These latter eight observers then switched tasks, and we measure cross-task transfer in the following section (Experiment 2). We do not include post-switch performance in this results section.

Figure 3A displays the accuracy results (fraction correct) of Experiment 1 for array mean discrimination, as a function of session number. The different curves reflect different levels of difficulty in terms of mean orientation difference between the two arrays, ORDA, from  $\pm 2^\circ$  (hard, orange) to  $\pm 12^\circ$  (easy, dark blue). The data are best fit to logarithmic improvement curves. Average performance over these levels is shown by the black dashed line, in which  $\text{accuracy} = 0.72 + 0.05 \times \ln(\text{session number})$ ;  $R^2 = 0.93$ . The improvement with training from the first two sessions (0.75) to the last two sessions (0.83) is highly significant ( $p < 0.0001$ ). There is an increase in accuracy with session at every difficulty level, with improvement for the easiest level being the smallest, and improvement for the hardest being the latest.

The same data are plotted in Figure 3C as performance accuracy versus ORDA for each session. Increased performance with larger ORDA is seen in the upward trend of each curve, and improvement from session to session is seen in the upward/leftward shift from curve to curve. For example, taking 75% correct as a threshold, the ORDA at threshold level decreases with session number, as shown in the figure inset. The best fit logarithmic threshold decay has orientation difference,  $\text{ORDA} = 6.5 - 1.6 \times \ln(\text{session number})$ ;  $R^2 = 0.96$ .

Figure 3B displays RT decrease with learning for mean discrimination. RT is a function of task difficulty, that is, mean-orientation difference, ORDA, between the two arrays, seen by the different curves of Figure 3C. In addition, RT decreases with session number for every level of difficulty, as seen by the downward-going curves. Average RT over these levels as a function of session number is shown by the black dashed line, in which  $\text{RT (ms)} = 873 - 230 \times \ln(\text{session number})$ ;  $R^2 = 0.97$ . The decreased RT with training from the first two sessions (790 ms) to the last two sessions (392 ms) is highly significant ( $p < 0.001$ ). Decrease in RT together with increase in accuracy supports the conclusion that training-based improvement is not a speed-accuracy trade off.

Accuracy results (fraction correct) of Experiment 1 for outlier detection as a function of session number is displayed in Figure 4A. The different curves reflect different levels of difficulty in terms of orientation difference between the outlier and the mean array orientation (and hence the edge of the array orientation range), ORDO, from  $\pm 15^\circ$  (hard, green) to  $\pm 30^\circ$  (easy, red). The data are best fit to logarithmic improvement

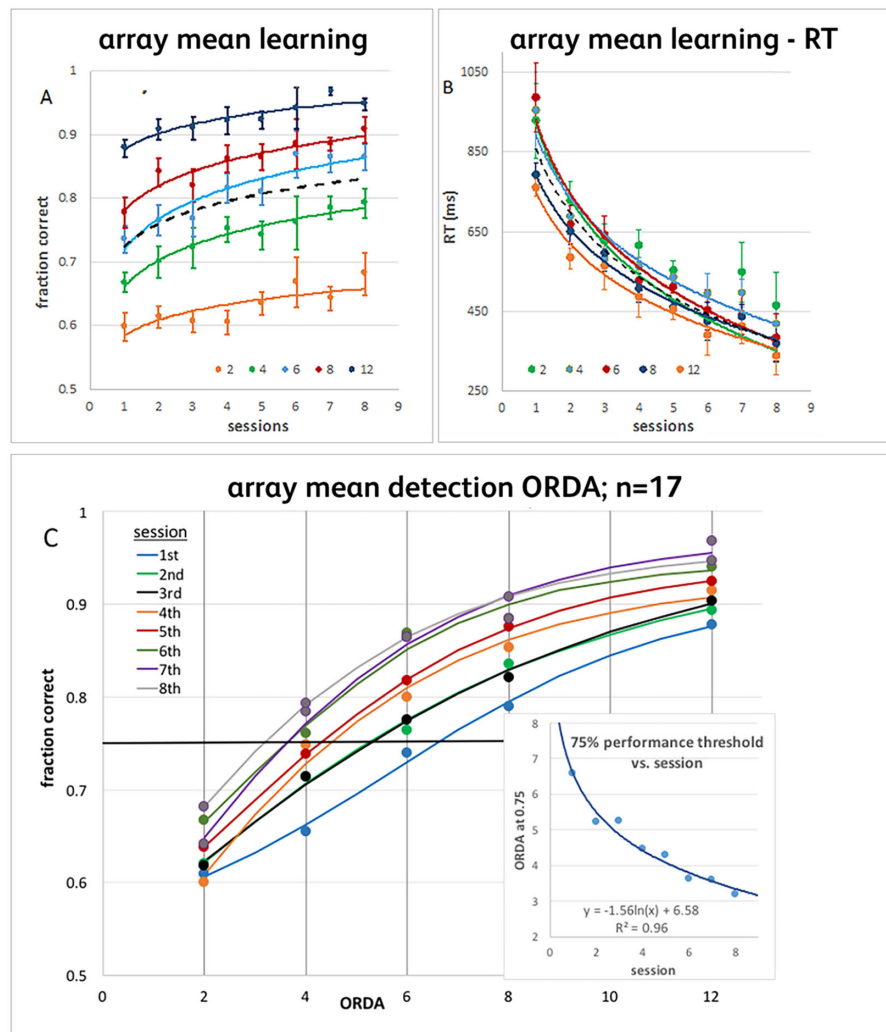


Figure 3. Perceptual learning in the mean discrimination task: Experiment 1;  $n = 17$ . (A) Performance accuracy as a function of session number, for various mean orientation differences between the two sequentially presented arrays (ORDA), from small difference ( $\pm 2^\circ$ ) to large difference ( $\pm 12^\circ$ ). Improved accuracy is seen at all levels of difficulty. Curves are best fit logarithmic functions. Dashed black line is average over difficulty level (see text for best fit equation). (B) RT decrease as function of session number, reflecting perceptual learning, at various difficulty levels (ORDA). (C) Performance accuracy as a function of task difficulty (ORDA) for each session, from first (light blue) to last (gray). Curves are best fit sigmoidal functions. Inset: ORDA at 75% correct as function of session number.

curves. Average performance over these levels are shown by the black dashed line, in which accuracy  $= 0.69 + 0.038 \times \ln(\text{session number})$ ;  $R^2 = 0.95$ . The improvement with training from the first two sessions (0.71) to the last two sessions (0.78) is highly significant ( $p < 0.001$ ). There is an increase in accuracy with session at every difficulty level, with greatest improvement for the first to second session.

The same data are plotted in Figure 4C as performance accuracy versus outlier orientation difference, ORDO, for each session. Increased performance with larger ORDO is seen in the upward trend of each curve, and improvement from session to session is seen in the leftward/upward shift from curve to curve. Taking 75% correct as a threshold, the ORDO

at threshold level decreases with session number, as shown in the figure inset. The best fit logarithmic threshold decay has orientation difference from array extreme orientation,  $\text{ORDO-VAR} = 14.3 - 2.1 \times \ln(\text{session number})$ ;  $R^2 = 0.90$ . The decreased RT with training from the first two sessions (805 ms) to the last two sessions (560 ms) is highly significant ( $p < 0.001$ ).

Figure 4B displays RT decrease with learning for outlier detection. Here too, RT is a function of task difficulty, that is, orientation difference between outlier and its array, ORDA, seen by the different curves of Figure 4C. In addition, RT decreases with session number for each level of difficulty, as seen by the downward-going curves. Average RT over these levels as a function of session number is shown by the

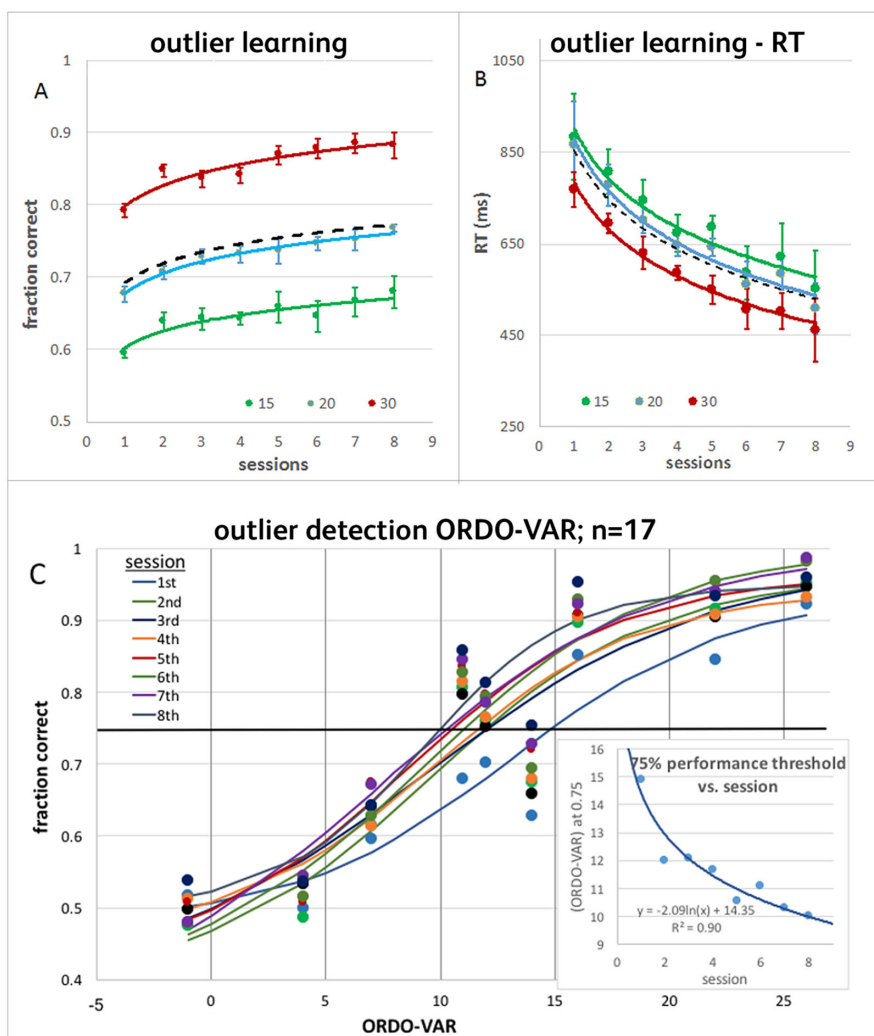


Figure 4. Perceptual learning in the outlier detection task: Experiment 1;  $n = 17$ . (A) Performance accuracy as a function of session number for three orientation differences (ORDO) between that of the outlier and the array mean orientation from small difference ( $\pm 15^\circ$ ) to large difference ( $\pm 30^\circ$ ). Improved accuracy is seen at all levels of difficulty. Curves are best fit logarithmic functions. Dashed black line is average over difficulty level (see text for best fit equation). (B) RT decrease as function of session number, reflecting perceptual learning, at various difficulty levels (ORDO). (C) Performance accuracy as a function of task difficulty measured here as distance between outlier orientation and that of the nearest array orientation (array orientation range edge, i.e., ORDO-VAR) for each session from first (light blue) to eighth (gray). Curves are best fit sigmoidal functions. Inset: (ORDO-VAR) at 75% correct as function of session number.

black dashed line, in which  $RT \text{ (ms)} = 855 - 160 \times \ln(\text{session number})$ ;  $R^2 = 0.98$ . Decrease in RT together with increase in accuracy supports the conclusion that training-based improvement is not a speed-accuracy trade off.

### Experiment 2. Ensemble perception learning transfer across tasks

One of the most essential issues in all learning is the issue of transfer of learned skills or information to new situations or circumstances (Druckman & Bjork,

1994). If we learn to drive in New York, will we be proficient in Los Angeles? If we learn on a Toyota, will we be capable with a Mazda? If we learn to drive with our right eye closed, will we be able to drive with our left eye closed? If we learn to stop quickly when there is a red traffic light over the road, will we stop quickly when the light is to one side? Or, to ask an educational question, if law students learn to respect other students' rights and to help those in need, will they respect and help during their careers?

The same issue is essential in the study of perceptual learning. If we learn to detect speedily a pop out target in the left visual field, will this learning transfer to the right visual field? If we learn to detect a right diagonal

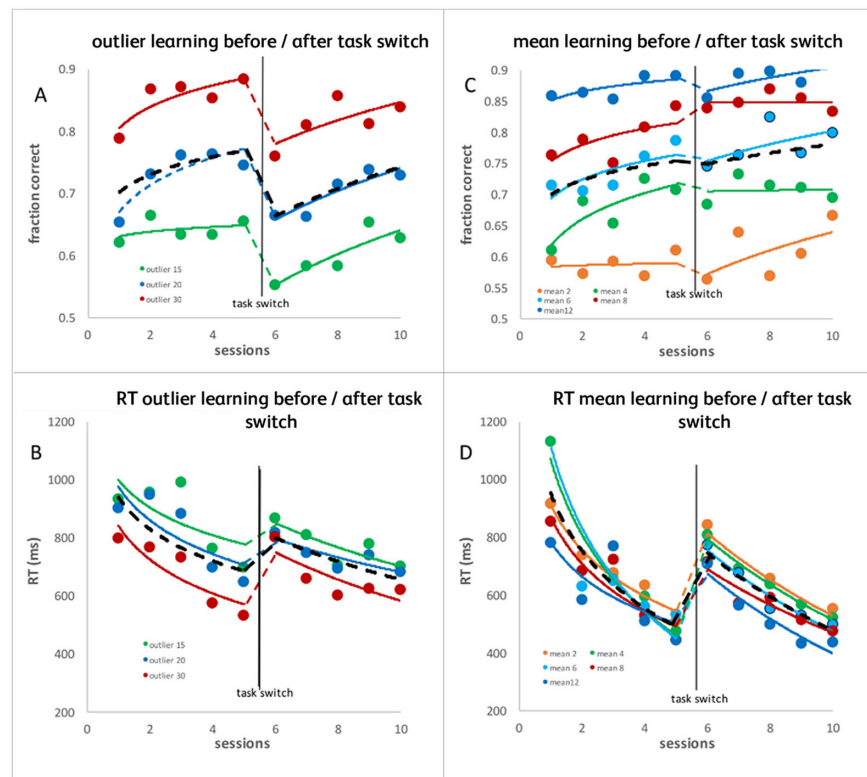


Figure 5. Perceptual learning transfer between tasks: Experiment 2. Participants ( $n = 4$ ) performed 5 sessions of outlier detection (A,B left half) and then 5 sessions of array mean orientation discrimination (C,D right half). Another set of participants ( $n = 4$ ) performed first the mean discrimination task (C,D left) and then the outlier detection task (A,B right). There is transfer of learning for the first set of participants, from outlier detection practice to mean discrimination performance, as seen in the high mean discrimination accuracy (C right) in these participants compared to the naïve mean discrimination participants (C left), as well as their low outlier detection reaction time (D right) compared to that of naïve participants (D left). On the other hand, there is no transfer, and even some interference so that outlier detection performance after mean discrimination practice (A right) is below the level of naïve participants (A left); nevertheless, there is some shortening of reaction time (B right) compared to naïve level (B left).

line, will this transfer to left diagonals? If we train with the right eye, will improvement transfer to performance with the left eye? All these have been dealt with at length in the perceptual learning literature and in other articles in the current special issue (see Introduction).

A most interesting type of transfer is transfer between tasks. Is acquaintance with an experimental situation and an experimental stimulus sufficient for improvement so that if we learn one task, improvement will also be seen with a second task—performed in the same environment and with the same stimuli? This issue has been studied less but is at the core of understanding perception and perceptual learning. Ahissar and Hochstein (1993) studied this issue using rectangular arrays of oriented bars, with participants reporting either global array rectangle orientation (wide or high), or local presence/absence of an outlier element. Training led to considerable improvement in the performed task, but nearly no improvement in the task that was not performed (see also Treisman, Vieira, & Hayes, 1992; Shiu & Pashler, 1992). There was a slight asymmetry in that performing the global rectangle

orientation task led to a small improvement in local outlier detection, which task was presumably performed implicitly and automatically during performance of the global task.

Eight MTurk observers participated in the current Experiment 2; four performed five sessions (every 2 days) of the mean discrimination task, followed by five sessions (every 2 days) of the outlier detection task, and four began by performing five sessions (every 2 days) of the outlier detection task followed by five sessions (every 2 days) of the mean discrimination task. Task protocols were like those of Experiment 1, except that the current observers only performed one task each day.

Results for these eight observers are shown in Figure 5, in which we plot data for each outlier orientation difference (ORDO: Figure 5A,B) or each array orientation difference (ORDA: Figure 5C,D). For the first five sessions, results are like those of the averaged data (the dashed lines) of Figures 3A and 4A, as they should be (the only difference is that in the present case the same observers will be tested in the



alternate task in their sessions 6–10; in fact, [Figures 3 and 4](#) include these observers). Transfer across tasks is seen in the performance following the first five sessions, when the task is switched between participants: those who performed mean discrimination now perform outlier detection, and those who were trained on outlier detection are now tested on mean discrimination.

The result is not symmetric. Rather, mean orientation discrimination accuracy following outlier detection training ([Figure 5C](#), sessions 6–10: 0.77) looks like a continuation of performance level towards the end of mean discrimination training, rather than a repeat of learning from scratch (sessions 1–5: 0.73; t-test:  $p < 0.001$ ). Similarly, there is speeding seen in the reaction time for mean discrimination following outlier detection learning ([Figure 5D](#): sessions 6–10: average 603 ms), compared to naïve performance (sessions 1–5: average 678 ms;  $p < 0.01$ ). Thus, there is transfer from practice with outlier detection to mean discrimination. This may result from the need to perceive the mean and range of the array in order to detect an outlier. Thus, practice with outlier detection might include implicit practice with determining array mean.

However, outlier detection accuracy following mean discrimination practice (sessions 6–10: 0.71) seems below outlier detection without any practice (sessions 1–5: 0.74)—rather than transfer enhancement, there is an interference effect ( $p < 0.001$ ). There is no transfer from mean discrimination training to outlier detection performance, perhaps because mean discrimination is an entirely global task, and the presence of orientation variability within each array reduces the automatic pop out effect of the outlier orientation (compare [Ahissar & Hochstein, 1993](#)). Looking for the array means to discriminate between them emphasizes mean without much attention to range, so observers do not automatically learn outlier detection. Nevertheless, there is some generalized improvement in RT in this case too (sessions 1–5: 788 ms vs. sessions 6–10: 723 ms;  $p < 0.05$ ).

[Ahissar and Hochstein \(1993\)](#) found that there was some learning of the orientation outlier detection, pop out task when its presence was implicit, and observers were performing the array global orientation task. The current results might be different from those due to our use of heterogeneous orientations within the arrays compared with the homogeneous orientations in the [Ahissar and Hochstein \(1993\)](#) presentations.

### Experiment 3. Ensemble perception implicit learning

Following suggestions that perceptual learning may take place even without attention, that is, when participants are performing a different task, we

tested for implicit performance improvement in our experimental paradigm.

We trained eight MTurk participants on one task, when the stimulus for the second task was present, but irrelevant to performance of the assigned task. Each observer participated in 17 sessions over 6 weeks, each Monday, Wednesday, Friday, as follows. Four observers (two men, two women) did five sessions of array mean discrimination, judging which of two successively presented arrays had a more clockwise mean orientation, in which the base orientation was  $60^\circ$  or  $70^\circ$ , and the difference in mean orientation between the arrays was parameter ORDA,  $\pm 2^\circ$ ,  $\pm 4^\circ$ ,  $\pm 6^\circ$ ,  $\pm 8^\circ$ , or  $\pm 12^\circ$ . One of the two arrays also had a bar with an outlier orientation in one of 12 locations, with an orientation difference, ORDO, from the mean of that array of  $\pm 15^\circ$ ,  $\pm 20^\circ$ , or  $\pm 30^\circ$ . Despite the presence of the uninformative outlier, observers were instructed to respond only according to the mean orientation of the arrays. Following completion of five sessions, we asked these participants to switch to nine sessions of the outlier detection task. Again, stimulus arrays had different mean orientations and one had an outlier, with the same stimulus parameters as in the first five sessions. However, now the observers were asked to perform the alternate task, that is, to detect the outlier, and array mean orientation was irrelevant. Finally, the participants were instructed to switch back to perform mean orientation discrimination for another three sessions.

Four other participants (three male, one female) performed the complementary tasks. They began with five sessions of the outlier detection task, whereas the two arrays irrelevantly had different mean orientations. Then they switched to nine sessions of the mean orientation task, and presence of the outlier was irrelevant. Finally, they performed three more sessions of outlier detection.

It is important to note that the irrelevant task could not help participants' performance because it was random, and thus irrelevant to the task. When performance was of the mean discrimination task, the outlier could appear in the more or the less clockwise oriented array equally, and when performance was of the outlier detection task, the more clockwise oriented array could be that with or without the outlier. Outlier orientation difference (ORDO) was measured from the mean of the array in which it was present and was either more clockwise or more counterclockwise than the mean.

The results are demonstrated separately for performance of the mean discrimination task ([Figure 6](#) top, bottom) and for performance of the outlier detection ([Figure 7](#) top, bottom). [Figures 6 and 7](#) top show performance accuracy, and [Figures 6 and 7](#) bottom show performance RT. To enable comparison of performance, in each graph we mix performance

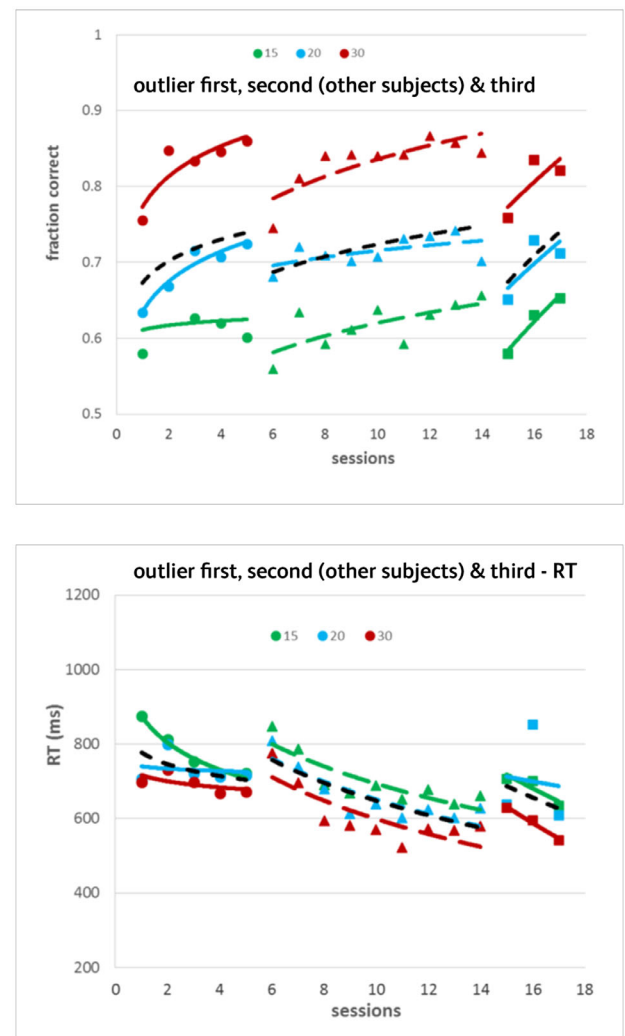
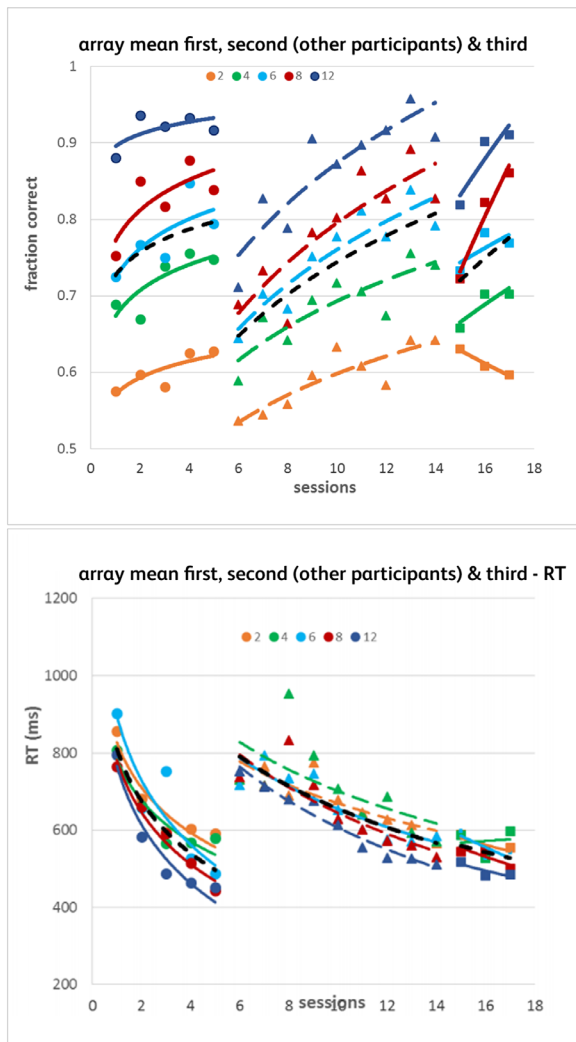


Figure 6. Implicit perceptual learning test: Experiment 3. Top: array mean orientation task performance accuracy as a function of session number for two groups of participants. In all cases, the two arrays presented differed in mean array orientation and one array had an outlier; participants were instructed which task to perform. Participant group 1 ( $n = 4$ ) performed the mean discrimination task for five sessions (left data points), followed by nine sessions of outlier detection (not shown), and then five more sessions of mean discrimination (right data points). Group 2 ( $n = 4$ ) performed outlier detection (not shown) followed by nine sessions of mean discrimination (shown middle data points). Comparing performance for these groups in their own first mean discrimination sessions (plotted data 1–5 vs. 6–10) demonstrates consistent interference by outlier practice to mean discrimination, that is, poorer performance for group 2 who had implicit exposure to arrays of different means. Similarly, group 1 exposure to outlier detection in sessions 6 to 14, interfered with their continued good performance after learning in sessions one to five. Bottom: RT for same participant groups and sessions. Interference effects are absent.

Figure 7. Implicit perceptual learning test: Experiment 3. Top, outlier detection task performance accuracy as a function of session number for two groups of participants. As in Figure 6, the two arrays presented differed in mean array orientation and one array had an outlier; participants were instructed which task to perform. Participant group 1 performed the outlier detection task for five sessions (left data points), followed by nine sessions of mean discrimination (not shown), and then five more sessions of outlier detection (right data points). Group 2 performed mean discrimination (not shown) followed by nine sessions of outlier detection (middle data points). Comparing performance for these groups in their own first outlier detection sessions (plotted data 1–5 vs. 6–10) does not show interference by mean discrimination practice to outlier detection, that is, here performance for group 2 who had implicit exposure to arrays of different means is like that of the naive group. Group 1 exposure to outlier detection in sessions 6 to 14 interfered with their continued good performance after learning in sessions 1 to 5. Bottom: RT for same groups and sessions. Interference effects are absent.

from the two groups of participants for the same task.

Figure 6 top shows performance of mean discrimination for the first five sessions and last three sessions of the group who performed mean discrimination first (and last), as well as the mean discrimination performance by the other group who performed this task in their sixth to 14th sessions. Performance after switching task (sessions 6–10: 0.70) is poorer than initially (sessions 1–5: 0.77 for the other group;  $t$ -test:  $p < 0.0001$ ). There is significant interference for performing and learning mean discrimination after considerable practice with the same stimuli and performing the outlier detection task. During performance of the first five sessions of outlier detection, the differences in mean orientation between the two arrays was irrelevant, and perception of this difference seems to have been suppressed actively (although subconsciously) by the participants. This suppression led to reduced mean discrimination performance in sessions six to 14, unlike the case in Experiment 2, Figure 5, in which there was no implicit presence of the secondary task elements. In addition, when switching back and forth, the group who began with mean discrimination had to do some more learning after having performed outlier detection in the middle, that is, performance was significantly lowered (sessions 15–17: 0.75 vs. sessions 3–5: 0.80;  $p < 0.001$ ). Thus the interference seems to be even for remembering the learned mean discrimination skill when performing a different task, outlier detection, in the middle.

Figure 6 bottom shows performance RT for mean discrimination. Performance, and even learning, after the switch is slower than initial learning, reflecting the interference found earlier ( $t$ -test: mean discrimination RT: sessions 1–5: 624 ms vs. sessions 6–10: 720 ms;  $p < 0.001$ ;) However, initial RT speeding due to initial training is fully maintained after back and forth switch ( $t$ -test: sessions 3–5: 548 ms vs. sessions 15–17: 543 ms;  $p = 0.38$  n.s.).

The differences in development and transfer measured by accuracy versus by RT suggests that there are different processes that determine these two performance parameters. Perhaps RT improvement is generalized across tasks, whereas accuracy requires specific learning, and presence of a confusing parameter leads to its being ignored or inhibited.

Figure 7 top shows performance of outlier detection for the first five sessions (average accuracy 0.73) and last three sessions of the group who performed outlier detection first (and last), as well as the outlier detection performance by the other group, who performed this task in their sixth to 14th sessions. There is little to no difference in performance when comparing performance after switching (0.72) with initial performance (0.73). In other words, there is little if any advantage of having performed mean discrimination for five sessions, even

with presence of the outlier (session 1–5 vs. 6–10:  $t$ -test:  $p = 0.31$ ). This may be interpreted in two ways: either there is little improvement and little interference from mean discrimination to outlier detection, or there may be both some learning and some interference and they largely cancel each other out. Looking at performance for the last three sessions, after switching back and forth (0.71), there is carryover of performance improvement from the original training (sessions 3–5;  $p = 0.12$ ). Performance of the mean discrimination task in the middle does not interfere much with outlier detection performance.

Figure 7 bottom shows performance RT for outlier detection. Learning after the switch (sessions 6–10: 692 ms) is faster than initial learning (sessions 1–5: 727 ms;  $p < 0.01$ ), and initial learning is fully maintained after back and forth switch (sessions 3–5: 709 ms vs. sessions 15–17: 657 ms;  $p < 0.05$ ), although five sessions may not have sufficed for full training. Again, the differences in accuracy versus RT measures suggests that RT is generalized, whereas accuracy is specific.

These results suggest that performance with an implicit second task might be like performance without the second task present at all. However, presence of the second implicit task seems to prevent perceptual learning transfer from task to task, suggesting it is suppressed when present but not performed. Furthermore, although second task presence leaves learning of the original task undisrupted, its performance is disturbed by switching tasks back and forth.

#### Experiment 4. Ensemble perception implicit learning parameter dependence

There is another aspect to consider when introducing implicit presence of the second task stimulus. When performing the mean discrimination task and looking for the more clockwise-oriented array, a clockwise turned outlier, when present on the more clockwise-oriented array, might make that array appear somewhat more clockwise on average (even though it is only one bar in 69). In this case, its presence would assist mean discrimination performance. However, if the clockwise-turned outlier appeared on the more counterclockwise array, it might make the task more difficult. Similarly, when looking for an outlier, if a counterclockwise-turned outlier appears on the more clockwise array, it may pop out more (perhaps because the array-to-array difference in most bars is in one direction, and that of the outlier bar is in the opposite direction). If, however, the outlier is turned clockwise and appears on the more clockwise array, (or counterclockwise on the counterclockwise array), it may be more difficult to detect it because all bars are

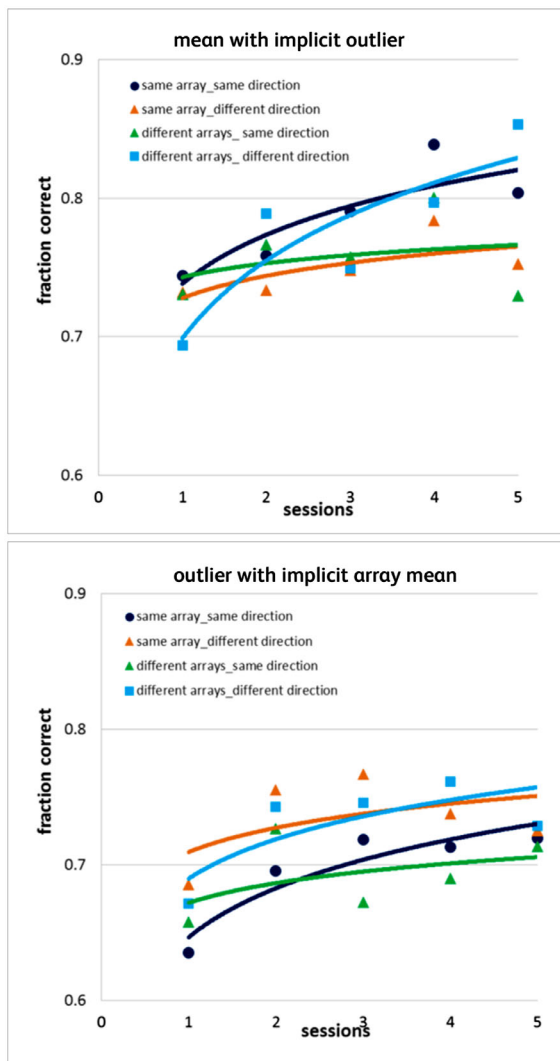


Figure 8. Impact of implicit presence of alternate task's features. The perceptual learning data of the first five sessions of Figures 6 and 7 top are divided between cases in which the clockwise-oriented mean and the outlier are on the same array or different arrays. Top:  $n = 4$ . Mean discrimination performance is improved when the outlier increases the difference between the mean orientations of the two arrays (dark and light blue), and is hindered when the outlier decreases this difference (orange and green). Bottom:  $n = 4$ . Outlier detection is improved when the outlier is nearer the horizontal (orange and light blue) rather than the vertical (dark blue and green), which is closer to the array mean.

on average more clockwise (or counterclockwise) than those on the other array.

To test this hypothesis, we reanalyzed the results for the first five sessions of Experiment 3. In Figure 8 top we plot performance for the mean discrimination task when there was an implicit outlier present, separating four cases: (1) when the outlier appeared on the array that was to be perceived and reported as more

clockwise, and the outlier was turned in the same direction as the array, that is, clockwise (dark blue); (2) when the outlier appeared on the array that was to be perceived and reported as more clockwise, and the outlier was turned in the opposite direction as the array, that is, counterclockwise (orange); (3) when the outlier appeared on the other array that was less clockwise and thus not to be selected for report, and the outlier was turned in the same direction as the array, that is, counterclockwise (light blue); and (4) when the outlier appeared on the other array that was less clockwise, and thus not to be selected for report, and the outlier was turned in the opposite direction as the array, that is, clockwise (green).

Mean detection performance is best when the outlier is turned clockwise, in the same direction as the array (dark blue), and, after the first session, also when the outlier is turned counterclockwise, in the same direction as the (nonreported) array (light blue). Performance is lower when the outlier is on the same clockwise-turned array, but itself turned in the opposite, that is, counterclockwise direction (orange), or is on the not-to-be-reported counterclockwise-turned array but itself is turned in the opposite direction, that is, clockwise (green).

Perhaps, when looking for the mean, if one stimulus bar (the outlier) is even more clockwise than the already more clockwise array, it adds something to the overall clockwise mean orientation of the array, so it is easier to see that this array is more clockwise. However, if the outlier stimulus bar is less clockwise in the generally clockwise array, then it interferes with discriminating this array as being more clockwise. Performance is therefore lower. The same effects are seen when the outlier is on the counterclockwise-turned array.

In Figure 8 bottom we plot performance for outlier detection when there was an implicit mean array orientation difference present, separating four cases: (1) when the to-be-detected outlier was turned clockwise and it appeared on the array that had a more clockwise mean orientation (dark blue); (2) when the to-be-detected outlier was turned counterclockwise and it appeared on the array that had a more clockwise mean orientation (orange); (3) when the to-be-detected outlier was turned clockwise and it appeared on the array that was less clockwise (light blue); and (4) when the outlier was turned counterclockwise and appeared on the array that was also less clockwise (green).

Unlike the results for mean discrimination (Figure 8 top), performance is best when the outlier is turned counterclockwise, whether on the clockwise- or counterclockwise-turned array (orange and light blue). Performance is lowest when the outlier is turned in the clockwise direction, whether on the clockwise- or counterclockwise-turned array (dark blue and green). This seems strange. We offer the following possible explanation: recall that  $0^\circ$  is horizontal and

90° is vertical, so that bars of orientation between 0° and 45° are categorically more horizontal, and those between 45° and 90° (and even 135°) more vertical. We hypothesize that detecting an outlier is easier when it is turned (15°, 20°, or 30° relative to mean array orientation) in the negative, counterclockwise direction rendering the outlier orientation closer to horizontal rather than closer to vertical, whereas the array bars are generally more vertical (mean 60° or 70°). This interpretation predicts that this effect should be exaggerated for cases of base mean orientation 60° (counterclockwise outlier 45°, 40°, or 30°) rather than 70° (55°, 50°, or 40°). This prediction was confirmed (effect was 23% compared with 12%).

## Summary and discussion

In a series of experiments, we found several significant perceptual learning effects. In all experiments, observers were presented with two arrays of stimulus bars with a distribution of orientations. The two arrays had different average orientations, and/or one array contained a bar with an outlier orientation (Figure 2). The task was to discriminate mean array bar orientation, or to detect the outlier. We found gradual and significant perceptual learning in all cases, demonstrated by improved performance accuracy and reduced RT, as well as a reduced orientation difference required to reach 75% performance threshold (Figures 3 and 4). This was true whether one task alone was trained, or if the two tasks were performed in interleaved fashion, with little, if any difference between these cases. This result joins similar findings of task performance improvement for many perceptual tasks (see Introduction).

Testing for cross-task transfer of learning, we switched the task requested of the observers following training with one of the tasks over five sessions. An asymmetry was found (Figure 5A,C). There is considerable transfer from training outlier detection to mean discrimination performance, and none in the opposite direction. There is even interference from mean discrimination training to outlier detection performance after the switch. As suggested earlier (Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018), detecting an outlier requires determining the edges or range of the observed distribution, which is closely related to the distribution mean. If you know the range limits, you are one step ahead to determine the middle of the range or its mean. Mean and range perceptions may be intimately linked. However, looking for the mean does not require perceiving outliers. Indeed, it has been reported that observers disregard outliers in computing the mean (Haberman & Whitney, 2010), so outlier perception may be actively obstructed. It may

therefore be difficult to switch to the second task that requires attending to just the element that, as found by Haberman and Whitney (2010), has been inhibited for five sessions.

Response speeding was found to transfer across tasks in both directions (Figure 5B,D). This result suggests that RT reduction might be a generalized phenomenon. Speeding might be related to stimulation structure, rather than being specific to the perceptual task being performed.

The third experiment tested for implicit learning. Will observers improve more for tasks whose stimulus elements are present, even when they are not asked to perform this task. In this experiment, all trials contained both an array that had a more clockwise mean orientation and an array that had an outlier, although these effects could be on different arrays, presented first or second. Again, the results are asymmetric. Surprisingly, experience with mean discrimination stimuli (while performing outlier detection) did not improve post-switch mean discrimination performance, and instead there was considerable performance interference (Figure 6). Observers seemed to actively (although implicitly) disregard or inhibit detection of the second task's features (mean orientation), so that this irrelevant feature does not disturb performance of the required task. Following five sessions inhibiting these irrelevant features, it becomes difficult for observers to switch and suddenly attend just to these features. Support for this interpretation comes from the difficulty found even for subsequent reverse switch, in which the original task is performed again, following nine sessions of the second task. All improvement gained by training five original sessions is lost during the nine intermediate sessions in which perception of the original mean discrimination task features must be suppressed.

However, outlier detection performance shows neither improvement nor interference (or perhaps both) from outlier presence during prior mean orientation discrimination performance (Figure 7). Still, the learning-based improvement due to original outlier detection training is lost when there are many intervening sessions of mean detection (with outlier presence). Interestingly, this interference too is present in performance accuracy but not in performance speed. RT measures are not reverted to original levels, suggesting that there are separate mechanisms for performance accuracy and speed, the one being specific and the other generalized.

A deeper look at the results of this third experiment, in which elements of both tasks were always present, led to further conclusions concerning implicit perception of task-irrelevant features (Figure 8). Mean discrimination was improved when the outlier orientation increased the difference between the mean orientations of the two arrays, and mean discrimination was disturbed

when the outlier orientation reduced the difference between the means. In contrast, outlier detection was improved when the outlier orientation differed from the array in the negative orientation direction, perhaps because this moved the bar from the more vertical to the more horizontal quadrant. Thus only in the mean discrimination task was the nonperformed task feature perceived to some extent and improved or disturbed performance when synergistic or antagonistic to the task being performed.

We conclude that perception and perceptual learning are intricate complex results of task and stimulation procedures. Features that seem irrelevant may lead to complex learning attributes, inducing improved performance, and learning transfer, or on the contrary, to interference of performance and/or learning transfer.

Several studies of ensemble perception have suggested that there is a relationship between perceiving ensemble statistics, mean and range or variance, and detecting set outliers. It has been suggested that one of the important advantages of ensemble perception is its enabling detection of outliers. However, two opposite relationships have been proposed. On the one hand, detecting outliers is important because these abnormal elements may be important. See, for example, the studies by Anne Treisman and her colleagues and the many studies in her wake, in which the goal of the task presented to participants was visual search for the outlier (Treisman & Gelade, 1980). However, it has been noted that recognizing an outlier is important to *not* take this unusual element into account when finding the mean and range or variance of the set.

## Conclusions

We have suggested that there is a relationship between ensemble and outlier perceptions because they share underlying mechanisms. A population code as suggested by Georgopoulos and colleagues (1986) would derive both the mean and the range of the set, without explicit representation of the identity of each individual element. Nevertheless, mean and range are two computations, and the visual system could easily perform one and not the other and certainly form a higher-level representation of one and not the other. The current results suggest that these two computations are not necessarily performed in tandem. Instead, when the task at hand directs participant concentration to mean perception, participants do not perform outlier detection in parallel, and there is no perceptual learning of this second task. However, performance of outlier detection, involving computation of the range edge, may include automatic computation of the set mean, leading to cross-task transfer in this case. Nevertheless, when

the array means are different, and the task is to detect an outlier in one of them, participants seem to actively disregard the different means, and when subsequently requested to perform mean discrimination, it is difficult for participants to start paying attention to this feature that was previously suppressed. When performing mean discrimination, however, and an outlier is present, because it is important to not include the outlier in the mean computation, it is therefore important to detect said outlier. Thus, having practiced outlier detection in order to exclude outliers from mean computation, subsequent performance of outlier detection is not inhibited and may even be a bit better than without prior implicit experience.

*Keywords:* ensemble perception, outlier detection, perceptual learning, pop out

## Acknowledgments

The authors thank Dov Sagi, Merav Ahissar, and many participants at the 2017 Perceptual Learning Workshop for constructive advice about this report. The authors thank Nachum Soroker and Yoram Bonneh for helpful discussions. The authors thank Yuri Maximov for help with programing, interaction with MTurk participants, and data analyses. This study was supported by the Israel Science Foundation (ISF).

Commercial relationships: none.

Corresponding author: Shaul Hochstein.

Email: shaul.hochstein@gmail.com.

Address: ELSC Safra Brain Research Center and Life Sciences Institute, Hebrew University, Jerusalem, Israel.

## References

- Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), 5718–5722.
- Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research*, 36(21), 3487–3500.
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401–406.
- Ahissar, M., & Hochstein, S. (2000). The spread of attention and learning in feature search: Effects of target distribution and task difficulty. *Vision Research*, 40(10–12), 1349–1364.

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8, 457–464.
- Ahissar, M., Laiwand, R., & Hochstein, S. (2001). Attentional demands following perceptual skill training. *Psychological Science*, 12(1), 56–62.
- Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K.B. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, 9, 1558–1564.
- Ahissar, M., Nahum, M., Nelken, A., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B*, 364, 285–299, <http://doi.org/10.1098/rstb.2008.0253>.
- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, 101, 34–40.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 19, 392–398, <https://doi.org/10.1111%2Fj.1467-9280.2008.02098.x>.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7345–7350, <https://doi.org/10.1073/pnas.0808981106>.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157–162.
- Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, 218(4573), 697–698.
- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6), 953–996.
- Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception & Psychophysics*, 71(3), 655–665, doi:[10.3758/APP.71.3.655](https://doi.org/10.3758/APP.71.3.655).
- Bauer, B. (2015). A selective summary of visual averaging research and issues up to 2000. *Journal of Vision*, 15(4):14, 1–15, <https://doi.org/10.1167/15.4.14>. [PubMed] [Article]
- Braun, J., & Sagi, D. (1991). Texture-based tasks are little affected by second tasks requiring peripheral or central attentive fixation. *Perception*, 20(4), 483–500.
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, 5, 10415.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, 20, 324–335.
- Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, 138, 289–301.
- Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, 237, 400–402.
- Druckman, D., & Bjork, R. A. (1994). *Learning, remembering, believing: Enhancing human performance*. Washington, DC: National Academy Press.
- Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.
- Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research*, 21(7), 1149–1151, 1153–1158.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585–612.
- Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, 14(9), 22, doi:[10.1167/14.9.22](https://doi.org/10.1167/14.9.22).
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771), 1416–1419.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17, 751–753.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734.
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception & Psychophysics*, 72(7), 1825–1838.
- Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe, & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York: Oxford University Press.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791–804.
- Hochstein, S., Khayat, N., Pavlovskaya, M., Bonneh, Y., Soroker, N., & Fusi, S. (2019). Perceiving category set statistics on-the-fly open access. *Journal of Vision*, 19, 225, doi:[10.1167/19.10.225a](https://doi.org/10.1167/19.10.225a).

- Hochstein, S., Pavlovskaya, M., Bonneh, Y., & Soroker, N. (2018). Comparing set summary statistics and outlier pop out in vision. *Journal of Vision*, 18(13):12, 1–13, doi:10.1167/18.13.12.
- Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience*, 10, 647–658.
- Hubert-Wallander, B., & Boynton, G.M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, 15(4):5, doi:10.1167/15.4.5.
- Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2007). Inverting houses and textures: Investigating the characteristics of learned inversion effects. *Vision Research*, 47(27), 3350–3359.
- Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, 88(11), 4966–4970.
- Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443), 250–252.
- Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, 18(9):23. doi: 10.1167/18.9.23.
- Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception & Psychophysics*, 81, 2850–2872, doi: 10.3758/s13414-019-01792-7.
- Lu, H., Qian, N., & Liu, Z. (2004). Learning motion discrimination with suppressed MT. *Vision Research*, 44(15), 1817–1825.
- Lu, Z. L., Yu, C., Sagi, D., Watanabe, T., & Levi, D. (2009). Perceptual learning: Functions, mechanisms, and applications. *Vision Research*, 49(21), 2531.
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11):9.1–9.8, <https://doi.org/10.1167/8.11.9>. [PubMed] [Article]
- Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, 128(1), 56–63, <https://doi.org/10.1016/j.cognition.2013.03.006>.
- Ofen, N., Moran, A., & Sagi, D. (2007). Effects of trial repetition in texture discrimination. *Vision Research*, 47(8), 1094–1102.
- Pavlovskaya, M., & Hochstein, S. (2011). Perceptual learning transfer between hemispheres and tasks for easy and hard feature search conditions. *Journal of Vision*, 11(1):8, doi:10.1167/11.1.8.
- Pavlovskaya, M., Soroker, N., Bonneh, Y., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, 27(7), 1397–1411.
- Rubenstein, B. S., & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A*, 7(9), 1632–1643.
- Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, 51, 1552–1566.
- Sale, A., Berardi, N., & Maffei, L. (2009). Enrich the environment to empower the brain. *Trends in Neurosciences*, 32(4), 233–239.
- Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, 11(1), 53–60.
- Seitz, A. R., & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature*, 422(6927), 3, <https://doi.org/10.1038/422036a>.
- Shiu, L.-P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, 52, 582–588.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14):19, 1–16, <https://doi.org/10.1167/10.14.19>. [PubMed] [Article]
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337.
- Thompson, B., & Liu, Z. (2006). Learning motion discrimination with suppressed and un-suppressed MT. *Vision Research*, 46(13), 2110–212.
- Treisman, A. (2009). Attention: Theoretical and psychological perspectives. In M. S. Gazzaniga, E. Bizzi, L. M. Chalupa, S. T. Grafton, T. F. Heatherton, & C. Koch, ... B. A. Wandell (Eds.), *The cognitive neurosciences* (pp. 189–204). Cambridge, MA: Massachusetts Institute of Technology.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology*, 105(2), 341–362.
- Tsushima, Y., Seitz, A. R., & Watanabe, T. (2008). Task-irrelevant learning occurs only when the



- irrelevant feature is weak. *Current Biology*, 18(12), R516–R517.
- Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858), 844–848.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129.
- Yamanashi-Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, 7, 13186.
- Zhaoping, L. (2009). Perceptual learning of pop-out and the primary visual cortex. *Learning & Perception*, 1(1), 135–146.