

Comparing set summary statistics and outlier pop out in vision

Shaul Hochstein

ELSC Safra Center for Brain Research, Hebrew University,
Jerusalem, Israel



Marina Pavlovskaya

Loewenstein Rehabilitation Center, Raanana, Israel
Sackler Faculty of Medicine, Tel Aviv University,
Tel Aviv, Israel



Yoram S. Bonneh

School of Optometry & Vision Science,
Mina & Everard Goodman Faculty of Life Sciences,
Bar Ilan University, Ramat-Gan, Israel



Nachum Soroker

Loewenstein Rehabilitation Center, Raanana, Israel
Sackler Faculty of Medicine, Tel Aviv University,
Tel Aviv, Israel



Visual scenes are too complex to perceive immediately in all their details. Two strategies (among others) have been suggested as providing shortcuts for evaluating scene gist before its details: (a) Scene summary statistics provide average values that often suffice for judging sets of objects and acting in their environment. Set summary perception spans simple/complex dimensions (circle size, face emotion), various statistics (mean, variance, range), and separate statistics for discernible sets. (b) Related to set summary perception is detecting outliers from sets, called “pop out,” which allows rapid perception of presence and properties of unusual, and thus, possibly salient items in the scene. To understand better visual system mechanisms underlying these two set-related perceptual phenomena, we now study their properties and the relationship between them. We present observers with two clouds of bars with distributed orientations and ask them to discriminate their mean orientations, reporting which cloud is oriented more clockwise, on average. In the second experiment, the two clouds had the same mean orientation, but one had a bar with an outlier orientation, which observers detected. We find that cloud mean orientation discrimination depends mainly on the difference in means, whereas outlier detection depends mainly on the distance of the outlier orientation from the edge of the cloud orientation range. Neither percept depends largely on the range of the set itself. A unified model of a population-code mechanism underlying these two phenomena is discussed.

Introduction

Gestalt psychologists pointed out the importance of grouping similar scene elements into sets (Koffka, 1935). Like categorization, perceiving set statistics allows responding to various elements belonging to the same class, as if they were identical. Relating to object groups as ensembles expands processing, attention, and memory limits (Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016; Utochkin, 2016) and facilitates comparing discernable groups (Chong & Treisman, 2003). Rather than encoding and retaining characteristics of every element, the brain perceives average set parameters and “discards information about individual items” (p. 160; Ariely, 2001).

Recent studies find that observers rapidly perceive set mean values for multiple object features, including size (Ariely, 2001; Chong & Treisman, 2003, 2005; Alvarez, & Oliva, 2008; Corbett & Oriet, 2011; Bauer, 2015; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015; Pavlovskaya, Soroker, Bonneh, & Hochstein, 2015; Khayat & Hochstein, 2018), orientation (Dakin & Watt, 1997; Khayat & Hochstein, 2018), brightness (Bauer, 2015; Khayat & Hochstein, 2018), position (Alvarez & Oliva, 2008), face identity, gender, emotional expression, or eye-gaze (de Fockert, & Wolfenstein, 2009; Haberman & Whitney, 2007) and general lifelikeness (Yamanashi Leib et al., 2015; for recent reviews see Haberman & Whitney, 2012; Bauer, 2015; Hochstein et al., 2015; Cohen et al., 2016). Furthermore, it has been shown that observers can compare

Citation: 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, <https://doi.org/10.1167/18.13.12>.

<https://doi.org/10.1167/18.13.12>

Received May 10, 2018; published December 18, 2018

ISSN 1534-7362 Copyright 2018 The Authors



the means of two sets of items seen simultaneously or sequentially (Chong & Treisman, 2005). Set statistics perception has also been reported for the auditory domain (McDermott, Schemitsch, & Simoncelli, 2013; Nelken & de Cheveigné, 2013), suggesting this is a basic, widespread, phenomenon.

Observer reports of set statistics include set feature variance or range, as shown directly (Morgan, Chubb, & Solomon, 2008; Solomon, 2010; Haberman, Lee & Whitney, 2015) and as evidenced by their reporting presence or absence of items depending on their being within or outside the range, respectively (Pollard, 1984; Dakin & Watt 1997). Recently, Khayat and Hochstein (2018) reported that observers perceive not only set mean but also set range, implicitly, automatically, and on-the-fly, on a trial-by-trial basis. Consequently, when asked to choose which of two stimuli had been included in a previously viewed sequence of elements, they easily reject those that are outside the set range. This finding, that when observers try to remember the identity of viewed elements and, instead, perceive the mean and range of the elements seen, lends strong support to the conjecture that they view all the elements globally, rather than focus on a few (see Discussion).

Since observers perceive set range, they should also be able to directly detect and identify outliers which deviate from a set. Indeed, numerous studies have shown that observers are very quick to notice a deviant item within a set of homogeneous items (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989; Ahissar & Hochstein, 1993; Wolfe, 1994) and succeed even when the set is heterogeneous (Duncan & Humphreys, 1989; Rosenholtz, 1999; Hershler & Hochstein, 2005). Inclusion of group range in summary statistics may be essential for determining set membership versus outliers (Treisman & Gelade, 1980; Haberman & Whitney, 2012).

Regarding set mean perception, we now ask if discriminating between sets with different means depends on the difference between their means alone, or if there is dependence also on the range of the sets. Does discriminating between sets depend on their ranges being separated and nonoverlapping? Is it much harder to determine the mean of a set with a broad range than a narrow-range set?

Regarding set outlier detection, we now ask what is the relationship, if any, between the two perceptual phenomena, set mean discrimination and outlier detection. Does the quick detection of deviant items—their popping-out of the set—depend on their distance from the mean of the set, on the breadth of the set range, on their distance from the edge of the range of the set, or on another, independent measure? Even if outlier detection depends on set range and not directly on outlier distance from the set mean, does this imply that perceiving set mean and set range are fully

independent? Furthermore, following Khayat and Hochstein (2018), who found implicit set mean and range perception for element sequences, we now ask if this finding may be extended to explicit perception, when all elements are presented at once, making it difficult to focus on set distribution extremes.

To answer these questions, we use a single experimental paradigm to test both set mean discrimination and set outlier detection. We shall show that these perceptual phenomena depend on different aspects of the set. Nevertheless, we suggest that set perception inherently includes both these different aspects, so that a single perceptual mechanism may be responsible for both perceptual phenomena.

Methods

Observers

As described below, observers participated either in the simultaneous-presentation or the successive-presentation experiment. Thirteen observers, students and coworkers at the Lowenstein Rehabilitation Center, participated in these experiments, nine in the first, simultaneous-presentation experiment (age, 42 ± 10 years; four women, five men), and four in the second successive-presentation experiment (age, 28 ± 3 ; one woman, three men), a sufficient number, as we shall see, for significant results. All had normal or corrected-to-normal vision. All but one (author MP of the simultaneous group) were naïve as to the goals of the study.

We also sent these experiments to two groups of Amazon Mechanical Turk participants (MTurks; 26 in the first simultaneous-presentation group and 21 in the second successive-presentation group). 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 our results with a much larger group of observers. We believe that there is benefit in combining a handful of in-house participants with a greater number of MTurks. We rejected about half of the MTurk participant data when there were indications that they were not performing the task, when accuracy rates were at 50% chance level even for the easy conditions, and/or reaction times were inappropriate (<100 ms or >3 s).

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

Experimental set-up

Stimuli were displayed on a 19-in. color CRT monitor controlled by dedicated OpenGL-based (Austin, TX) software running on a Windows PC. The video format was true color RGB, 100-Hz refresh rate, with a 1024×768 pixel resolution occupying a $14^\circ \times 11^\circ$ area. Luminance values were gamma corrected, and the mean luminance was ~ 30 cd/m². Sitting distance was 0.7 m, and experiments were administered in near darkness. MTurk programs were written using Adobe Flash. 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 also in successive mode was to control for subjects not fixating the central circle between the two simultaneously presented clouds.

Stimuli and procedure

After the observer initiated the trial by pressing the central mouse key, a single fixation circle appeared at the center of the monitor, with a diameter of 1.1° . Following a second central mouse keypress, the fixation circle disappeared and a pair of arrays of bars were presented, unmasked, for 100 ms (MTurks: 200 ms) to the left and right of fixation, as demonstrated schematically in Figure 1. Arrays were 6° in diameter, with the nearest edges 2° from fixation. Each array contained 69 dark bars arranged in a 9×9 grid excluding three in each corner. Bar positions were jittered by up to 0.2° to avoid array homogeneity. Bar length \times width was $0.7^\circ \times 0.05^\circ$ and bar orientation was 60° or 70° mean \pm a variation factor, VAR, which served as the first variable of the study; (90° 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 variation factor, VAR, was set to be $\pm 4^\circ$, $\pm 8^\circ$, $\pm 16^\circ$ or $\pm 32^\circ$. For VAR = 32° , there are 65 orientations (with 1° steps) in the range $(60^\circ - \text{VAR})$ to $(60^\circ + \text{VAR})$, i.e., 28° to 92° or $(70^\circ - \text{VAR})$ to $(70^\circ + \text{VAR})$, 38° to 102° , 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° (or 70°), in order not to change the mean. This random placement was done independently for the

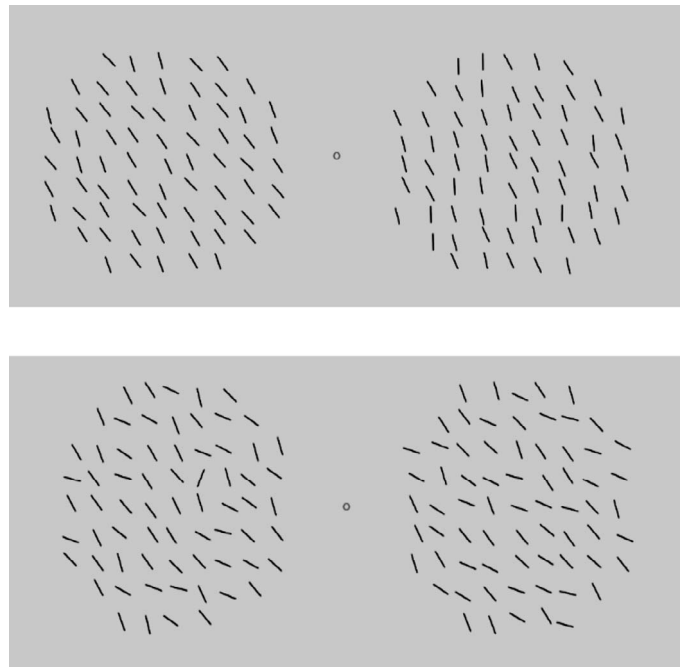


Figure 1. Participants viewed two circular arrays of bars of various orientations. Top: Mean orientation discrimination task; Observers reported which set mean is tilted more clockwise (in the example, the right one). 2nd row. Outlier detection task; Observers reported which set contains an outlier (the left one).

two arrays of the display. We only used VAR = 32° for the first group tested and found it to be too difficult. We therefore dropped this value for the remaining participant groups and do not report these results here.

In Experiment 1, the mean orientation discrimination test, 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. Following a pilot study to determine the range of interest, experimental sessions included trials with ORDA set to be 4° , 8° , or 12° , in random interleaved order; we added ORDA = 2° and 6° for in-house and 16° for Mturk participants. Observers were instructed to respond by clicking the left or right mouse button when perceiving the left or right array as having a more clockwise rotation, respectively.

In Experiment 2, the outlier detection test, the two arrays had the same mean orientation (again, randomly 60° or 70°), but one of the arrays had an outlier bar, which had an orientation which 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 5×5 positions, choosing 12, 16, 18, or 24 locations to be tested in each session; fewer locations allow more repetitions/location. Experimental sessions included trials with ORDO set to

be $\pm 15^\circ$, $\pm 20^\circ$, or $\pm 30^\circ$, in random interleaved order; (a pilot study including $\pm 7^\circ$ and $\pm 10^\circ$ determined this to be the range of interest). Observers clicked the left or right mouse button when detecting the outlier in the left or right array, respectively.

The same group of in-house participants performed both experiments in a series of sessions over several days. Each day's session included one experiment, either mean orientation discrimination (typically 200–520 trials/session; 10–26 trials/data point/ participant) or outlier detection (192–576 trials/session; 16–48 trials/data point/ participant), or both, one following the other, with a coffee break in between. The two experiments were interleaved over the days of testing, allowing us to measure between-session perceptual learning.

MTurk participants performed a single session of one experiment (120 trials orientation discrimination, 10 trials/data point/participant, or 216 trials outlier detection, 18 trials/data point/participant), so comparisons here are between participants, and we could not test perceptual learning.

We presented the second group of in-house and MTurk participants with the two arrays successively (in-house: 100 ms, MTurk: 150 ms, presentation; 300 ms interstimulus interval) in the center of the screen, instead of simultaneously side by side. They clicked the left mouse button to indicate the first array and the right mouse button for the second. Conditions were otherwise the same as for the first groups of in-house and MTurk observers, except that MTurk participants in the successive presentation group performed both experiments, sequentially, first a mean discrimination block (180 trials with ORDA = 0° , 4° , 6° , 8° , 12° , or 16° ; the first as a baseline test for bias, expecting, and finding, an average 50-50 response; 10 trials/data point/participant), and then an outlier detection block (216 trials; 10 trials/data point/participant).

In-house observers participated in 10 sessions, five for orientation discrimination and five for deviant detection (with each session including orientation discrimination: 400–600 trials, 10–15 trials/data point/participant; outlier detection: 576 trials, 12 or 24 outlier positions, 48 trials/data point). The MTurk groups performed both tasks in a single sitting, 120 trials for mean discrimination, (including 5 repetitions each of trials with 6 ORDAs, 4 VARs), and 216 trials for deviant detection (including 3 ORDOs, 4 VARs, and 18 outlier positions). In all cases, base mean orientation was randomly 60° or 70° .

Response measures

We measure response accuracy and reaction time (RT) for each experiment, each observer and each set of

parameters (VAR and ORDA or ORDO). Plots of results are presented against these variables or their difference (see Results) and statistical tests were performed, including *t* tests and Logistic regression analysis. Note that our experiments were of the 2-Alternative Forced Choice (2-AFC) paradigm so that every trial results either in a Hit plus Correct Rejection (CR) or in a Miss plus False Alarm (FA). Thus accuracy is both Hits/(Hits + Misses) and CR/(CR+FA).

We performed logistic regression analyses of the results to determine the relative effects of the two parameters, VAR and ORDA or ORDO.

Results

We first plot observer performance accuracy versus ORDA or ORDO, to see the impact of the difference in means of the two arrays or outlier difference from the array mean. Here VAR, the variance of the orientations of the arrays is a second variable, and we wish to determine if there is any dependence on this second variable. Figures 2 and 3 show the results for the four groups of observers, as follows: Figure 2 displays performance accuracy for the two experiments, for the four participant groups together, as well as results for each group separately, and Figure 3 shows performance Reaction Times (RT) for the two experiments. Each graph of Figures 2 and 3 display results for both mean discrimination (saturated colors) and outlier detection (lighter colors). The plots are generally nonoverlapping since significant performance was found for larger orientation differences for outlier detection (ORDO) than for mean orientation discrimination (ORDA).

This is the first finding of the comparison experiments: Outliers are detected only when they have large orientation differences from the array in which they are embedded. By comparison, array mean orientation is discriminated even for small orientation differences between two arrays. Thus, the range of above-chance performance for outlier detection is shifted to larger orientation differences than is the range for mean discrimination (15° – 30° vs. 2° – 16°). Comparing outlier detection and mean discrimination for similar ORDs, (15° and 12°), mean discrimination is more accurate (87% vs. 60%; $p < 0.01$), and it is more accurate even for the entire range tested (78% vs. 69%; $p < 0.005$). Similarly, mean discrimination is faster for similar ORDs (703 ms vs. 788 ms), though for the entire ranges tested they are similar (766 ms vs. 734 ms).

In all cases there is a monotonic increase in performance accuracy and monotonic decrease in RT with increasing orientation difference, between array

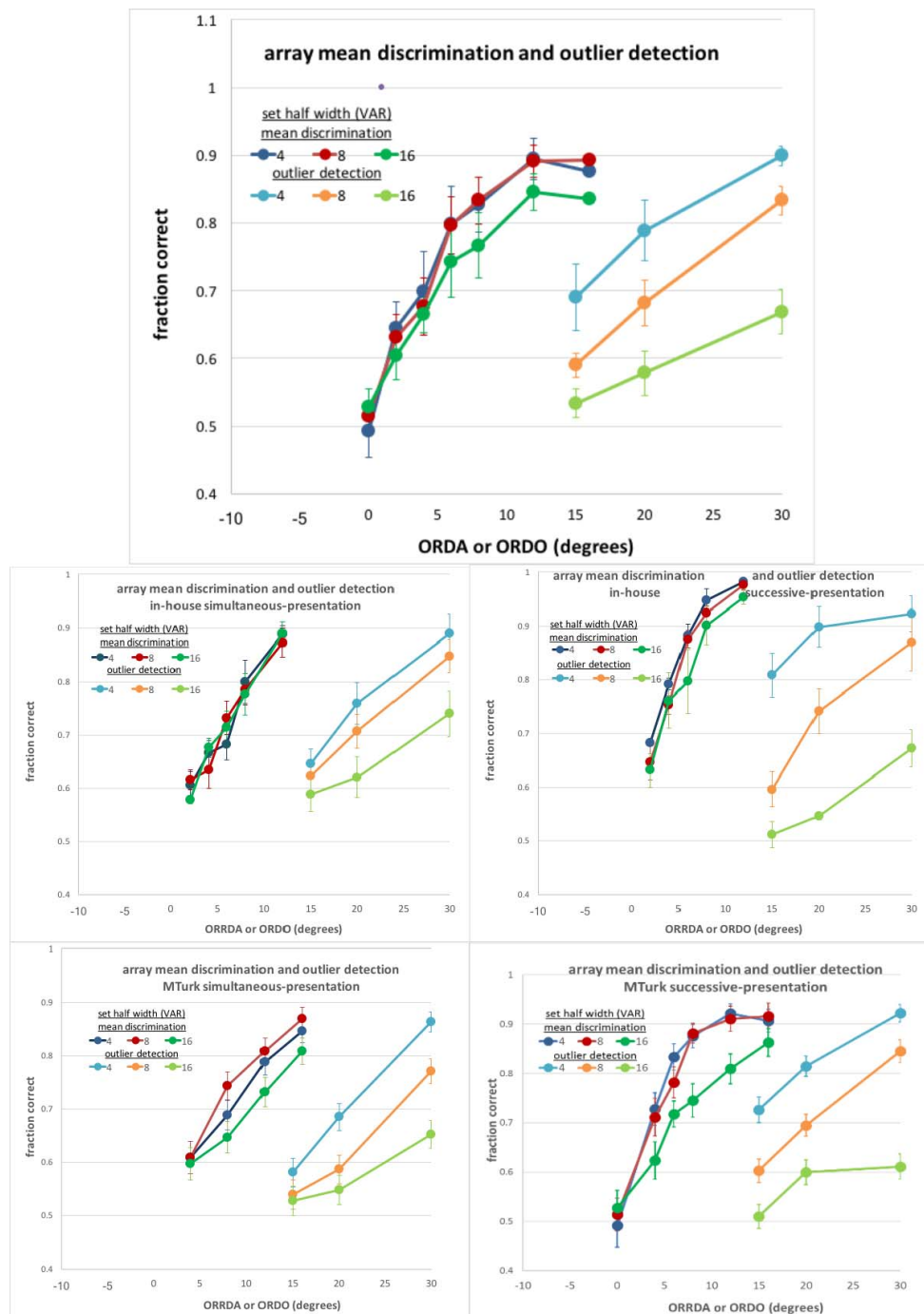


Figure 2. Performance accuracy dependence on orientation difference (ORDA and ORDO) average results for all observers (top). The lower graphs are separate results for in-house simultaneous (middle left) and successive (middle right) presentation, and for MTurk simultaneous (bottom left) and successive (bottom right) presentation. In each graph the left plots (saturated colors) are for performance on the array orientation discrimination task as a function of the difference between the mean orientations of the two arrays to be discriminated and the right plots (unsaturated colors) for performance on the outlier detection task, as a function of the difference between the outlier orientation and the mean orientation of the arrays. Performance improves in both cases with orientation difference from near chance, 0.5, to nearly 0.9. When there is no orientation difference (ORDA or ORDO = 0; tested only for MTurk successive presentation group), response is 50% chance, as it must be. Plots are shown for three array ranges, from 4° to 16° half range, VAR, in different colors. Note that there is nearly no dependence on array range for mean discrimination, but considerable dependence for outlier task accuracy (*t* test, *p* < 0.01). Error bars here and in subsequent figures are between participant standard errors.

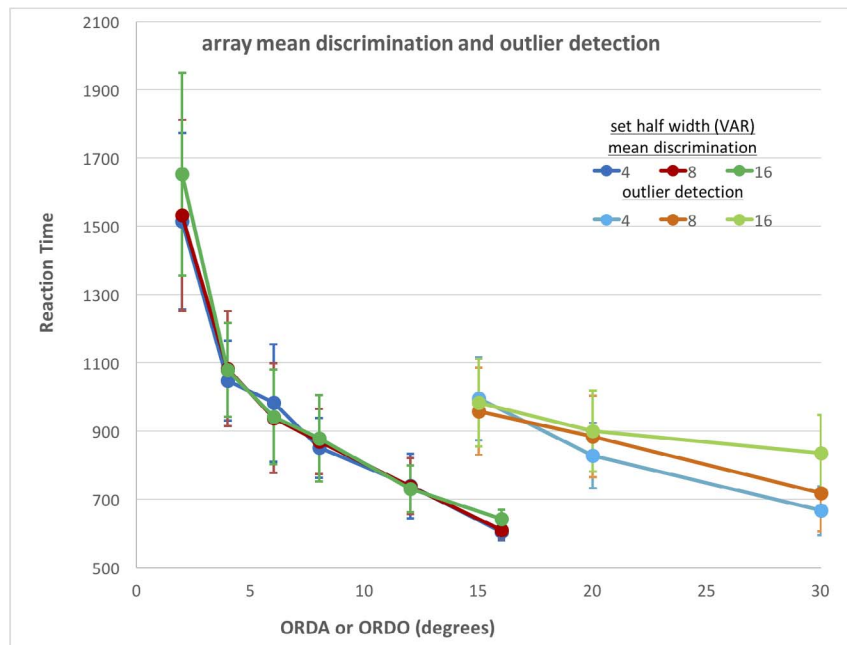


Figure 3. Reaction Times (RT) for the same conditions as Figure 2. Decreasing RTs indicate that the performance improvement is real and not simply a performance/reaction time tradeoff. No significant dependence on array range, especially, again, for the array orientation discrimination task (left plots).

means (ORDA) or between deviant and array (ORDO). The changes in both parameters, accuracy and RT, in opposite directions indicates that this is true improvement rather than a speed-accuracy tradeoff.

The second finding is that there is no dependence on array set range, or VAR, for mean discrimination accuracy (all between VAR t tests are nonsignificant). In contrast to this result, there is a large gradual dependence on this variable for outlier detection accuracy ($p < 0.05$ for all between VAR t tests; compare Figure 2, left and right plots). For the RT measure there is very little dependence on array range for either experiment.

All these conclusions are supported by comparing results for either simultaneously presented or successively presented arrays, and for in-house or MTurk participants.

The lower graphs of Figure 2 display the results for each of the four observer groups separately, middle row for in-house observers, and bottom row for MTurk observers. The graphs are similar, confirming the results. We found that the in-house participants were generally somewhat slower, but more accurate (t test accuracy: $p < 0.001$; RT: $p < 0.001$). Comparing in-house groups versus Mturk groups, in-house groups were more accurate in detecting outliers (71% vs. 64%; $p < 0.001$) but Mturk groups were considerably faster (610 vs. 931 ms; $p < 0.001$), and in-house groups were more accurate also in discriminating means (78% vs. 69%; $p < 0.001$) but Mturk groups were faster (684 vs.

1145 ms; $p < 0.001$), here, too. These between-group differences may reflect speed-accuracy tradeoffs.

In addition, the left graphs are for simultaneous presentation, and the right graphs are for successive presentation. Again, the graphs are quite similar. Here, performance is somewhat better for the successive presentation than for the simultaneous presentation, perhaps because in this case both arrays are presented near fixation.

For mean discrimination, comparing successive versus simultaneous presentation, successive presentation leads to significantly faster responses (668 ms vs. 865 ms; $p < 0.001$), both for the in-house (674 ms vs. 1046 ms; $p < 0.001$) and Mturk (662 ms vs. 684 ms; $p < 0.001$) groups, and performance with successive presentation is more accurate (84% vs. 73%; $p < 0.001$), both for the in-house (89% vs. 78%; $p < 0.001$) and Mturk (80% vs. 69%; $p < 0.001$) groups.

For outlier detection, comparing successive versus simultaneous presentation, performance for successive presentation is significantly faster (699 ms vs. 771 ms; t test, $p < 0.001$), but only for the in-house groups (776 ms vs. 931 ms; $p < 0.001$), not the Mturk groups (621 ms vs. 610 ms), and performance for successive presentation is more accurate (72% vs. 68%; $t < 0.07$), but only significantly so for the Mturk groups (70% vs. 64%; $p < 0.01$) not the in-house groups (73% vs. 71%; $p < 0.3$). Thus, performance is better for both groups, showing up for the in-house groups in speed and for the MTurk groups in accuracy.

Note that sets are discriminated even when the difference in mean orientation of the two sets is much smaller than the range of the sets, so that there is large overlap in the orientations present in the two sets. For example, with set half-range $VAR = 16^\circ$, and $ORDA = 8^\circ$, the ranges could be $44^\circ\text{--}76^\circ$ and $52^\circ\text{--}84^\circ$, respectively (or, with 70° mean, $54^\circ\text{--}86^\circ$ and $62^\circ\text{--}94^\circ$), and the mean orientations are 60° and 68° , respectively (or 70° and 78°). Thus, the mean orientation of each array is present in the other array, as well. All four observer groups are $\sim 75\%$ – 80% correct when judging which set mean is more clockwise despite the two means being both well within the ranges of both sets.

This is not the case for outlier detection. When $VAR = 16^\circ$, detection is well above chance only at $ORDO = 20^\circ$, that is when the outlier orientation is well outside the range of the array orientations.

We looked at observer results for response biases. Some observers were biased to respond right and others left, but these biases only appeared for the extremely difficult conditions where accuracy was near 50% chance in any case. There was considerable perceptual learning for these tasks, to be discussed separately, reflecting perceptual learning found for many rapid, seemingly automatic tasks (Ahissar & Hochstein, 1993; Karni & Sagi, 1993).

We sought the underlying cause of the differences between the results for mean discrimination and outlier detection. As we saw in Figure 2, mean discrimination depends on array mean difference but is nearly independent of set range, while outlier detection seems to have strong dependences on both variables, outlier orientation difference from the array mean as well as array range. In addition, the range of orientation difference for above-chance accuracy is much lower for mean discrimination than for outlier detection ($2^\circ\text{--}16^\circ$ vs. $15^\circ\text{--}30^\circ$). Perhaps, we asked, detection of an outlier depends on a single variable, namely the distance of the outlier from the edge of the range of the array. That is, as long as the outlier is within the array range, it doesn't pop-out because it is indistinguishable from the other members of the array. In order to be detected as an outlier, perhaps it must be significantly distant from the edge of the array.

To test this hypothesis, we plot outlier detection versus the orientation distance from the array edge, measured as the distance from the mean minus the set half-range. (Outlier orientation – set edge orientation = (outlier orientation – mean orientation) + (mean orientation – set edge orientation) = $(ORDO - \text{Set range}/2) = (ORDO - VAR)$).

On the right side of Figures 4 and 5 we plot the dependence of performance (Figure 4) and Reaction Time (Figure 5) for outlier detection as a function of the distance of the outlier orientation distance from the array edge ($ORDO - \text{Set Range}/2$). Note that the plots

for different ranges now coincide forming a single curve, confirming that outlier detection depends on this single variable, and not separately on the two variables, distance from mean and set range. For comparison, the left side plots of Figures 4 and 5 display the dependence of mean discrimination performance and RT on this same variable (difference between means – set range). Now the plots for different ranges are separated, suggesting that the correct dependence is only on the single variable $ORDA$, the difference between the array means irrespective of the set range, as shown in Figures 2 and 3.

As a statistical test of these two different dependences, we performed a logistic regression analysis for our data. The results are displayed in Table 1, and presented graphically in Figure 6. The assumption of logistic regression analysis is that performance accuracy depends on the two variables ($ORDA$ or $ORDO$ and set half-range) in the form of

Accuracy

$$= 1/[1 + \exp(a + b \times ORD + c \times VAR)] \quad (1)$$

where a , b , and c are constants or fit parameters. If the half-range, VAR , is not a significant determinant, we will expect that parameter c will be much smaller than parameter b . On the other hand, if the important factor is distance from the set edge, then we will expect parameters b and c to be of opposite sign and similar magnitude. These expectations are born out in Table 1. Figure 6 displays the results of the two tests as logistic regression plots, i.e., the line is exactly the above equation, with best fit parameters, and the points are the data themselves (saturated colors, mean discrimination; light colors, outlier detection).

Discussion

We tested participant performance in two tasks designed to use paradigms that are as close as possible. In both cases two arrays of lines of heterogeneous orientations were presented, either simultaneously, on two sides of the screen or sequentially in the same central screen position. The two tasks were to discriminate between the two array mean orientations and report which array contained lines that were on average more clockwise, or to detect which array contained a line with a deviant outlier orientation.

The most significant result was that these judgments were quite different. Outlier detection requires significantly larger difference between the orientation of the outlier and that of the array mean, than the between-array orientation difference required for mean orientation discrimination.

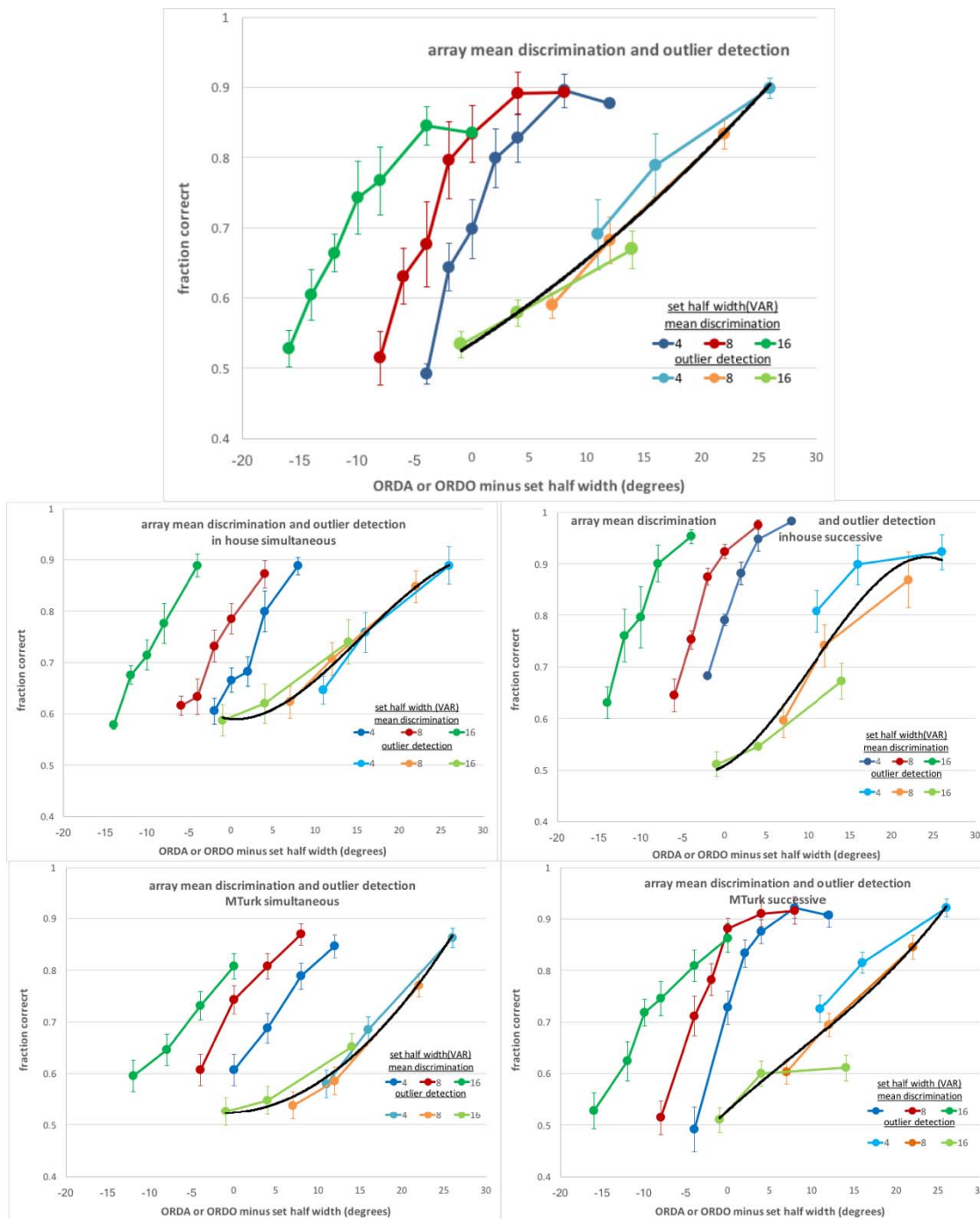


Figure 4. Performance accuracy dependence on the orientation distance from the set range edge. Results are shown when averaged for all observers (top) and separately for in-house observers with simultaneous (middle left) and successive (middle right) presentation and for MTurk observers with simultaneous (bottom left) and successive (bottom right) presentation. The abscissa is the distance from the set edge, equal to the difference between the orientation difference (ORDA or ORDO) and the set half-range (VAR), where the orientation difference is between the means of the arrays (ORDA for array discrimination, left plots in each graph) or between the deviant and the array mean (ORDO for deviant detection, right plots in each graph). These are the same data as Figure 2, replotted on the new abscissa. Plots shown for three VAR, or array half-ranges, from 4° to 16°. There is now nearly no dependence on array range for the outlier detection task, as indicated by the cumulative data black trend line, but considerable dependence for the mean discrimination task (t test, $p < 0.001$). Compare Figure 2.

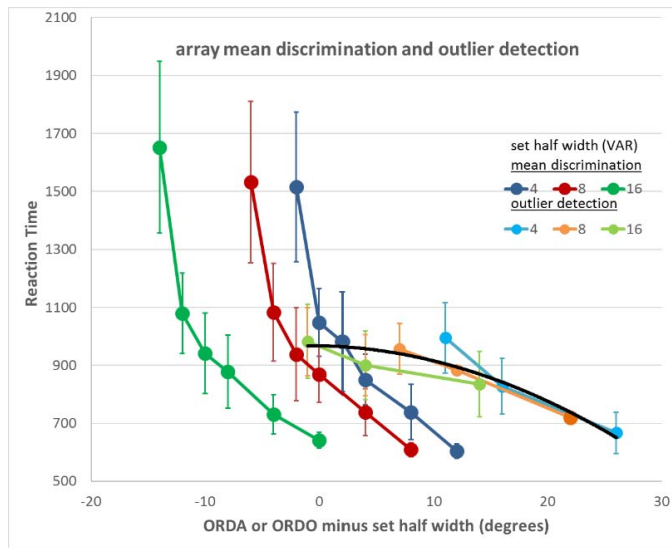


Figure 5. Reaction Time dependence on distance from array set range edge for all observers and both presentation conditions. Compare Figure 3.

Furthermore, different effects are seen when plotting the results as a function of orientation difference, either between the two arrays to be discriminated, or between the outlier and the array mean. For orientation mean discrimination, there is no dependence on array range. The only feature that determines mean discrimination is the orientation difference between the means. On the other hand, for outlier detection, the analogous orientation difference, that between the array mean and that of the deviant, does not satisfactorily describe the results. There remains an apparent dependence on array range, as well. This dependence on two factors is resolved as follows. Plotting the results as a function of the distance of the outlier orientation from that of the array edge, we now find congruence between the data for different array ranges. There is no remaining dependence on array range. Thus, the only determinant of outlier detection is the distance of the outlier orientation from the edge of the array, rather than from its mean.

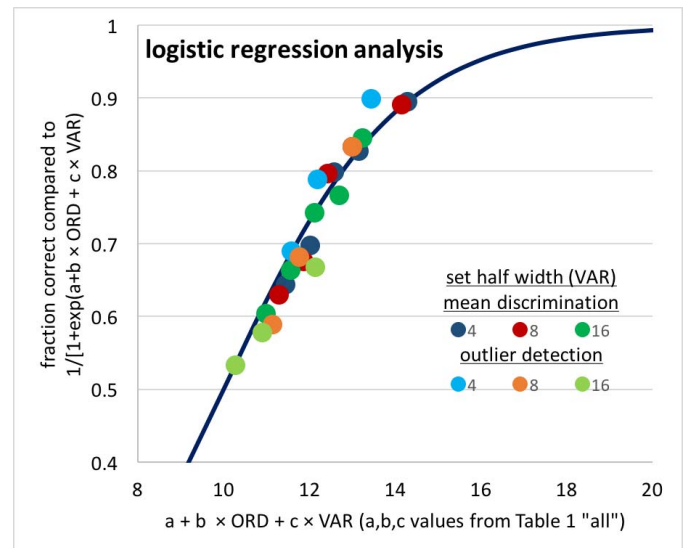


Figure 6. Logistic regression analysis for all subjects together. Data from Table 1 “all” results. Abscissa, arbitrary units to satisfy Equation 1 with parameters of Table 1; ordinate: fraction correct for same ORD and VAR values; black line is fit to Equation 1. ORD is respectively ORDO or ORDA.

These different dependences make sense, as follows. Set orientation mean is perceived for each array, and their means can be compared based on the precision of mean perception, which turns out to be independent of the variance or range of each array. Overlap of the two set ranges is irrelevant to this comparison. On the other hand, for a single element to be perceived as an outlier, it must be just that; it must lie outside the range of the set. Here, too, however, the dependence is only on distance from the range edge, and not on the size of the set range itself.

These results were confirmed by testing four groups of observers. Two groups of in-house participants under controlled environmental conditions, and two groups of online MTurk participants. The first in-house and first MTurk groups performed the tasks with simultaneous presentation on two sides of the monitor. The second in-house and MTurk groups were tested

Test	Observers	a	b	c	-c/b
Mean discrimination	9 in-house simultaneous	-0.345	-0.156	+0.006	0.04
	26 MTurk simultaneous	-0.6024	-0.1455	+0.0411	0.28
	4 in-house sequential	-0.5959	-0.2359	+0.0306	0.13
	21 MTurk sequential	-0.5029	-0.1077	+0.0299	0.28
	all	-0.5443	-0.1401	+0.0181	0.13
Outlier detection	9 in-house participants	-0.191	-0.068	+0.0587	0.86
	26 MTurk simultaneous	-0.5303	-0.0619	+0.0999	1.6
	4 in-house sequential	-0.6385	-0.0799	+0.1337	1.7
	21 MTurk sequential	-0.4552	-0.0603	+0.0938	1.6
	all	-0.2732	-0.0704	+0.0809	1.15

Table 1. Logistic regression parameters.

with successive presentation of the two arrays to be discriminated. The results of all four groups are similar, with only minor differences in speed or accuracy, but all showing the same essential effects described in the preceding paragraph.

Our results should be considered in light of contrasting findings suggesting that set mean perception and outlier “pop-out” accuracy decrease with set variance (e.g., see Dakin 2001; Rosenholtz, 2001; Haberman et al., 2015). For set mean, the type of elements being averaged may be important; line, circle, or face elements may be different than Gabor patterns, which lack clear borders and may cluster when crowded so that stimulus noise becomes critical. Regarding outlier detection, search among a few elements might be critically different than search in a large array or cloud of stimuli, and there may be a step change between homogeneous and heterogeneous distractor distributions. Further research is needed to distinguish between contrasting conditions.

The literature of set summary statistics regarding mean perception is quite broad and clear, though some have questioned whether set perception requires global perception of the entire set, suggesting that viewing a few may suffice (Myczek & Simons, 2008a, 2008b; but see also Ariely, 2008; Chong, Joo, Emmanouil, & Treisman, 2008), especially if the few include the largest and smallest. Dakin (2001) found that when controlling for element density (and thus for internal noise), the number of elements sampled to determine mean orientation, increases with the number presented, suggesting the number sampled might be all those present. Interestingly, it has been suggested that support for a mechanism integrating all elements without focused attention to any would come from finding set mean perception even when trying to recognize individual items (Myczek & Simon, 2008b). Such a supporting finding was provided in Khayat and Hochstein (2018). The current results also support this conclusion, since detecting the outlier depends on perceiving the range, and knowing the range depends on knowing all the elements (or at least the extreme ones).

In a related issue, there is question as to the type of mean perceived. Is it a simple arithmetic mean, are the elements perceived instead on a logarithmic scale, and is the computation of mean, median, or mode? Most studies are unable to answer these questions since they use small ranges of equally-sampled stimuli, so that there is little difference between these values. Nevertheless, the current understanding is that arithmetic mean is probably usually the best candidate (Bauer, 2015; but see, e.g., Zohary, Scase, & Braddick, 1996; Kimura, 2018).

Regarding the second statistic, there is even less information. Is it variance or range that is perceived?

Interestingly, Solomon (2010) asked observers to report which of two sets had the larger variance and found that estimates of orientation variance are more efficient and precise than estimates of mean orientation. Yet, he concludes that observers may be inferring variance from perceiving range. Detecting outliers, whether explicitly, as here, or implicitly as studied by Khayat and Hochstein (2018), would support the conclusion that it is range rather than variance that is detected, with quite clear determination of the edges of the range.

Having found that there is clear detection of the edges of the range of sets reintroduces the question raised above, namely, perhaps knowledge of the range is the basis for determination of the mean, which may be simply the average of the two extreme members of the set. Although Chong and Treisman (2005) presented various set distributions, these were all symmetric. Future experiments would need to present skewed set distributions to differentiate between these alternatives.

We conclude that outlier detection depends on outlier distance from the set edge. As mentioned, most studies of outlier detection—often termed “pop-out”—used homogeneous distractors, so that the distance from the mean and from the edge were equal. Still, some studies did use heterogeneous distractors (e.g., see Hershler & Hochstein, 2005), though not as a set. Duncan and Humphreys (1989) formulated a general principle whereby outlier search efficiency grows with target-distractor difference and with distractor-distractor similarity. This may be comparable with our own initial result that outlier detection increases with distance from set mean and decreases with enlarged range. It is more difficult to compare these results with our conclusion that these two parameters may be united into a single variable of outlier distance from set range edge, since their test elements were not along a continuum.

Recently, Haberman et al. (2015) tested explicit perception of facial expression variance. Their results fit a linearly growing error rate (reducing accuracy) with variance, but are also consistent with a fixed rate up to very large variances, especially for nonholistic stimuli (upside-down faces); we may not have reached the equivalently large range.

Does the above difference (dependence on distance between group means vs. dependence on distance from group’s range edge) imply separate computational mechanisms for the two perceptual tasks? We have suggested that this is not the case. Rather, observers may use a single computational mechanism for both tasks. It has been found that set summary statistics perception inherently includes both set mean and set range (Pollard, 1984; Dakin & Watt 1997; Solomon, 2010; Khayat & Hochstein, 2018). Thus, the bases for both mean discrimination and outlier detection are

inherently available as soon as a set is detected. We have suggested that the neural mechanism used is a population code (Georgopoulos, Schwartz, & Kettner, 1986) that encodes both the mean and the range of the stimulus set (Hochstein, 2016; Pavlovskaya, Soroker, Bonneh, & Hochstein, 2017a, 2017b). Recent evidence supports implicit perception of both mean and range, when participants are not asked to attend to either (Khayat & Hochstein, 2018).

The suggestion that a population code is used to determine set mean also answers the question of how the visual system computes mean set values without first knowing values for each element separately. Due to broad tuning and neuron receptive field overlap, a population code representation is necessarily used even for perceiving individual element values, as confirmed by adaptation phenomena (Georgopoulos et al., 1986). The same population code, with a broader range of neurons over space and time, might be responsible for perceiving mean values for sets of elements.

Interestingly, using a population code to determine the mean requires as a first step determination of the range of active neurons, to avoid the result being swamped by distant noise. This is easily done by including only responses that are well above the noise level. Thus, discriminating between array means and detecting an outlier both depend on the same population encoding, but on different aspects of this code: The set mean depends on population vector sum, and set outlier detection on population above-noise range. Interestingly, we found that mean discrimination is faster than outlier detection. This is especially poignant, recalling that outlier detection was Anne Treisman's original "preattentive" task (Treisman & Gelade, 1980), and is known to be very rapid (Ahissar & Hochstein, 1993).

Once the visual system reads out the edges of the set distribution, any element whose response curve is within the boundaries of the set distribution curve will be accepted as a set member to be included in the mean computation. Spurious deviances, which only result in a slightly broader envelope, may be accepted as set members. On the other hand, a truly exceptional element, whose response curve is represented as a separate peak beyond the outskirts of the envelope, will not be included in set mean computations, and will be recognized as an outlier. The position of this separate range of above-noise neuron responses will again be determined by a population code analysis.

We conclude that detection of a difference between two visual arrays depends mainly on the difference between their means, whereas detecting an outlier in an array depends mainly on the distance between the outlier and the edge of the array. In both cases, we find that performance depends only a little on the range itself of the array(s).

Set membership is determined by feature proximity; set range edges by feature-map response discontinuity. This new view provides a basis for feature search properties, explaining when set outliers "pop-out." Behavioral results support this relationship between cross-set comparison and outlier detection. We conclude that conscious perception begins with ensemble statistics, including outlier pop-out, only later focusing on individual set elements (Hochstein & Ahissar, 2002). Finally, note that categorization might use similar population encoding to represent category prototype (mean) and boundary (range; Goldstone, Kersten, & Carvalho, 2012).

Keywords: set summary statistics, pop out, population code

Acknowledgments

We thank Anne Treisman, of blessed memory, for discussions over many years on this and other subjects of mutual interest. Thanks to Robert Shapley, Howard Hock, Maya Bar Hillel, Merav Ahissar, and Ehud Zohary for insightful comments on earlier drafts of this paper. We thank Yuri Maximov for programming assistance and managing the interface with the Amazon online participants. This study was supported by a grant from the Israel Science Foundation (ISF).

Commercial relationships: none.

Corresponding author: Shaul Hochstein.

Email: shaulhochstein@gmail.com.

Address: ELSC Safra Center for Brain Research, 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*, 5718–5722.
- Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*, 1558–1564.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398.
- Ariely, D. (2001). Seeing sets: Representation by

- statistical properties. *Psychological Science*, *12*, 157–162.
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, *70*(7), 1325–1326.
- 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]
- Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.
- 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.
- Dakin, S. C. (2001) Information limit on the spatial integration of local orientation signals, *Journal of the Optical Society of America*, *18*(5), 1016–1026.
- Dakin, S. C., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*, 3181–3192.
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology: QJEP*, *62*, 1716–1722.
- Duncan, J., & Humphreys, G. W. (1989) Visual search and stimulus similarity, *Psychological Review*, *96*(3), 433–458.
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986, September 26). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416–1419.
- Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2012) Concepts and Categorization. In A.F. Healy & R.W. Proctor (Eds.), *Comprehensive handbook of psychology* (Vol. 4, pp. 607–630). New Jersey: Wiley.
- Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4):16, 1–11, <https://doi.org/10.1167/15.4.16>. [PubMed] [Article]
- 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. (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, NY: Oxford University Press.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707–1724.
- Hochstein, S. (2016) The power of populations: How the brain represents features and summary statistics. *Journal of Vision*, *16*(12):1117, <https://doi.org/10.1167/16.12.1117>. [Abstract]
- Hochstein, S., & Ahissar, M. (2002) View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* *36*, 791–804.
- Hochstein, S., Pavlovskaya, M., Bonneh, Y., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, *15*(4):7, 1–17, <https://doi.org/10.1167/15.4.7>. [PubMed] [Article]
- Karni, A., & Sagi D. (1993, September 16). 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, 1–14, <https://doi.org/10.1167/18.9.23>. [PubMed] [Article]
- Kimura, E. (2018) Averaging colors of multicolor mosaics, *Journal of the Optical Society of America. A, Optics, image science, and vision*, *35*, B43–B54.
- Koffka, K. (1935). *The principles of Gestalt psychology*. London: Routledge and Kegan Paul.
- McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*, 493–498.
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, <https://doi.org/10.1167/8.11.9>. [PubMed] [Article]
- Myczek, K., & Simons, D. J. (2008a) Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788.
- Myczek, K., & Simons, D. J. (2008b). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, *70*(7), 1335–1336
- Nelken, I., & de Cheveigné, A. (2013). An ear for statistics. *Nature Neuroscience*, *16*, 381–382.

- 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*, 1397–1411.
- Pavlovskaya, M., Soroker, N., Bonneh, Y., & Hochstein, S. (2017a). Statistical averaging and deviant detection in heterogeneous arrays. *40th European Conference on Visual Perception Abstracts*, *40*, 160.
- Pavlovskaya, M., Soroker, N., Bonneh, Y., & Hochstein, S. (2017b). Statistical averaging and deviant detection may share mechanisms. *2017 Neuroscience Meeting, Washington, DC: Society for Neuroscience*. 365:4.
- Pollard, P. (1984). Intuitive judgments of proportions, means, and variances. *Current Psychological Research & Reviews*, *3*, 5–18.
- Rosenholtz, R. (1999) A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*, 3157–3163.
- Rosenholtz, R. (2001) Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception & Performance*, *27*, 985–999
- 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]
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Utochkin, I. S. (2016) Visual enumeration of spatially overlapping subsets. *Russian Journal of Cognitive Science*, *3*, 4–20.
- Wolfe, J. M. (1994) Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989) Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, *15*(3), 419–433.
- Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2015) Fast ensemble of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*, 1106–1130.
- Zohary, E., Scase, M. O., & Braddick, O. J. (1996). Integration across directions in dynamic random dot displays: Vector summation or winner take all. *Vision Research*, *36*, 2321–2331.