NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

As a manuscript

Vladislav Khvostov

DISTRIBUTIONAL STATISTICS OF VISUAL FEATURES IN ENSEMBLE PERCEPTION AND CATEGORIZATION

Summary of the dissertation
for the purpose of obtaining academic degree
Doctor of Philosophy in Psychology

Academic supervisor:
Doctor of Science, Igor Utochkin

Moscow, 2022

Table of Contents

Three published articles were selected for the defense:

1. Khvostov, V.A., & Utochkin, I.S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision, 19* (8): 3, 1-18. DOI: 10.1167/19.9.3
2. Utochkin, I.S., Khvostov, V.A., & Stakina Y.M. (2018). Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions. *Cognition*, *179,* 178-191. DOI: 10.1016/j.cognition.2018.06.016
3. Utochkin, I.S., Khvostov, V.A., & Wolfe J.M. (2020). Categorical grouping is not required for guided conjunction search. *Journal of Vision, 20*(8): 30, 1-22. DOI: 10.1167/jov.20.8.30

The results are also published in the following article on this topic:

4. Khvostov, V.A., Lukashevich, A.O., & Utochkin, I.S. (2021). Spatially intermixed objects of different categories are parsed automatically. *Scientific Reports, 11:* 377, 1-8. DOI: 10.1038/s41598-020-79828-4

The dissertation was prepared at the Laboratory for Cognitive Research, National Research University Higher School of Economics.

1. Introduction

1.1. General research problem

Every second the visual system deals with lots of various objects in a scene. Even though severe limitations of focused attention and working memory (Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988) prevent detailed processing of all objects (Wolfe et al., 2011), observers typically experience no difficulties with seeing them all at the same time. One possible solution to this paradox of visual perception is the idea that the visual system represents some general statistical information about the whole set of objects without representing and storing information about each individual object (Alvarez, 2011; Ariely, 2001). It can be accomplished via computing statistical moments for feature distributions of a set of visible objects (Whitney & Yamanashi Leib, 2018) – ensemble summary statistics. It was shown that observers can extract the first and the second statistical moments which correspond to the average value of a visual feature (Alvarez & Oliva, 2008; Ariely, 2001; Bauer, 2009; Chong & Treisman, 2003, 2005b, 2005a) and its range/variance (Dakin & Watt, 1997; Morgan et al., 2008; Norman et al., 2015; Solomon et al., 2011). Also, observers can estimate the approximate number of objects in a scene without counting them one by one (Burr & Ross, 2008; Chong & Evans, 2011; Halberda, Sires, & Feigenson, 2006).

A broad spectrum of visual dimensions can be compressed into ensemble summaries: orientation (Alvarez & Oliva, 2009; Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), color (Gardelle & Summerfield, 2011; Maule & Franklin, 2015), size (Ariely, 2001; Chong & Treisman, 2003, 2005b, 2005a), motion (Watamaniuk & Duchon, 1992), even emotional expressions of faces (Haberman & Whitney, 2007) and many others (Florey, Clifford, Dakin, & Mareschal, 2016; Leib, Kosovicheva, & Whitney, 2016; Sweeny & Whitney, 2014). This provides a solid basis for guiding human behavior in various situations. Ensemble summary statistics can be represented perceptually rather than inferred "cognitively" which is supported by the evidence from adaptation aftereffects (Burr & Ross, 2008; Corbett et al., 2012; Norman et al., 2015; Ying & Xu, 2017). Ensemble summaries are extracted quickly (as rapidly as 50–200 ms, Chong & Treisman, 2003; Whiting & Oriet, 2011) and often with limited or absent conscious access to individual objects (Alvarez & Oliva, 2008; Ariely, 2001; Corbett & Oriet, 2011; Parkes et al., 2001).

Recent studies showed that the visual system can even represent the whole distribution of objects' features (Chetverikov et al., 2016, 2017a, 2017b, 2017c; Kim & Chong, 2020; Oriet & Hozempa, 2016). These findings suggest that simple summaries such as mean, variance, and numerosity are not the only things conveying ensemble information. Rather, it suggests that ensemble representations store quite rich information about the whole set of objects while mean, variance, and numerosity can be calculated based on these rich distributional representations (Khvostov et al., 2021).

How does the visual system use this rich information about the whole distribution of features? Can it be used for parallel and independent extraction of different ensemble summaries? What role does the distributional information play in everyday cognitive tasks such as visual search or rapid categorization of multiple objects? The present work summarizes research that my collaborators and I have done to investigate these questions.

1.2. Research goals

1) To test whether the calculations of different ensemble statistics can be done in parallel.

2) To study whether the calculations of different ensemble summaries are carried out by a common cognitive mechanism or independently, by several mechanisms.

3) To investigate the role of the shape of a feature distribution in the rapid visual categorization and segmentation of multiple objects defined by conjunctions of two features.

4) To study the role of the shape of a feature distribution in the conjunction and feature search tasks.

1.3. Methodology and theoretical basis of the work

The current work relies on and develops the following theoretical ideas:

1) sets of objects and scenes are represented in the compressed form of *ensemble summary statistics* (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003; Whitney & Yamanashi Leib, 2018).

2) the visual system uses the distributional characteristics of an ensemble for performing rapid categorization of multiple objects - *the theory of rapid visual ensemble-based categorization and segmentation* (Utochkin, 2015).

3) the visual perception process can be conceptually divided into "deep" and "shallow" processing, e.g. attention and preattention (Neisser, 1967; *Feature integration theory*: Treisman & Gelade, 1980), focused and distributed attention (Treisman, 2006), non-selective and selective pathways of scene and object processing (Wolfe et al., 2011).

4) the visual system uses the output from "shallow" processes (preattention) to efficiently guide "deep" processes (attention) – e.g., *Guided Search Theory* (Wolfe, 1994, 2021).

1.4. Hypotheses

1) The visual system extracts different statistical summaries from a set of objects independently (via distinct cognitive mechanisms) and in parallel (without any cost of dividing attention between the statistics).

2) The visual system uses the shape of the feature distributions as a cue for rapid visual categorization and segmentation of multiple objects defined by conjunctions of features along two sensory dimensions. The more feature dimensions have a distribution consisting of separate peaks ("segmentable" distribution), the more successful the categorization is.

3) The shape of the feature distributions influences the visual search speed: a presence of separate peaks in the distribution ("segmentable" distribution) will improve the search performance.

1.5. Research methods

All studies in this work are controlled psychophysical experiments conducted in laboratory settings. In various studies, different tasks and report paradigms were used: Study 1 used ensemble summary estimation with an adjustment method, Study 2 used a texture segmentation task with a two-alternative forced choice, and Study 3 used a visual search with speeded "yes-no" response. Statistical analysis was done in JASP software (JASP 0.9.0.1; JASP, Amsterdam, the Netherlands) and R. Statistical procedures included a series of repeated-measures ANOVA with subsequent pairwise t-tests and Pearson correlations.

1.6. Summary of scientific novelty

Most previous studies in the field of ensemble statistics were focused on investigating questions about extracting isolated statistical summaries (mostly, mean) along one feature (e.g., size) from a set of objects. The current work tests and develops a new idea that the visual system has access to much richer information about the whole feature distributions of different features rather than just individual statistical summaries. This rich information is used to drive performance on different perceptual tasks.

1) We investigated how the visual system deals with situations where it needs to extract several statistical summaries for a set of objects at the same time. We asked two important questions: (1) Does the visual system extract different summaries via independent cognitive mechanisms or a single mechanism? (2) Can the visual system extract several summaries in parallel (without the cost of dividing attention between these summaries)? We introduced a new paradigm, a version of the dual task implemented for the ensemble calculation which allowed us to test these two tightly connected questions in one task. Also, we developed a new way of data analysis for the dual-task trials based on within-individual correlation. It can directly test whether errors in reporting two statistics from the same trial are correlated which makes the analysis more powerful than in previous paradigms. Using these methods, we discovered new strong evidence that the visual system extracts several summaries independently and in parallel.

2) Previous works showed that the feature distribution along a single sensory dimension influences performance in cognitive tasks with multiple objects (e.g., Chetverikov et al., 2016; Utochkin & Yurevich, 2016; Im et al., 2021). For the first time, we tested how the visual system uses distributions of several features for such tasks. We discovered the conditions when the visual system can use several feature dimensions simultaneously for the rapid visual categorization of multiple objects defined by a conjunction of these features. Our studies revealed that the successful categorization can happen only when the distributions of both features have several separable peaks: e.g., objects are only big and small without any intermediate sizes. We proposed a new mechanism that explains these results and provided empirical evidence for this mechanism.

3) For the first time, we studied how the feature distributions along two sensory dimensions affect the deployment of attention over a visual search for a conjunctively defined target. Our study revealed conditions where they have no effect and where they strongly modulate the efficiency and speed of visual search. We obtained new evidence that the visual system can implement efficient conjunction search even in the absence of separable peaks within distractor distributions, i.e., in the absence of distractor grouping. We discovered that the Guided Search model (Wolfe, 1994, 2021) can perfectly explain these results. Therefore, our work establishes a new important connection between the theory of rapid visual ensemble-based categorization and segmentation (Utochkin, 2015) and seminal theories of visual search – e.g., the Guided Search model (Wolfe, 1994, 2021). Using the Guided search model, we predicted a very counterintuitive result: with certain distribution characteristics, visual search along one feature dimension can be much slower and more ineffective than visual search along two feature dimensions. For the first time, we tested and confirmed this prediction.

1.7. Theoretical significance

The current work contributes to the development of a set of important modern theories of visual perception and attention, such as the theory of ensemble summary statistics (Alvarez, 2011; Whitney & Yamanashi Leib, 2018), theory of rapid visual categorization and segmentation (Utochkin, 2015), and Guided search model (Wolfe, 1994, 2021). It advances our understanding of the architecture of visual representations: what information about a set of objects is available for the visual system and how the system uses it.

1.8. Applied significance

The results of the present work can be used for optimizing the way the information is visualized and displayed: e.g., how one should mark different objects to be sure that people easily can segment one group from another (even if the objects from different groups are spatially intermixed) or find a target object among others. Also, ensemble summary statistics is a topic tightly connected to mathematical statistics, therefore, the results can be used to improve teaching mathematical statistics using more intuitive examples from visual statistics. Some results of this work are used in undergraduate courses, “Cognitive Psychology” and “Psychology and Neurophysiology of Perception and Attention”, at the HSE University.

1.9. Statements for the defense

1) Several ensemble summaries can be extracted from a set of objects in parallel: without the cost of dividing attention between these statistics. The read-out of different summary statistics is accomplished via partially distinct, uncorrelated cognitive mechanisms which do their calculations independently.

2) The visual system uses the shapes of feature distributions as cues for rapid visual categorization and segmentation of multiple objects defined by a conjunction of two features. The prerequisite of such categorization is the “segmentability” of both feature distributions (the presence of several separate peaks representing likely categories). A plausible mechanism for carrying out such categorization is the half-split strategy: to select the object group along one feature dimension and compare the average value along a second feature dimension.

3) The visual system bypasses the absence of the distributional “segmentability”

and effectively searches for a known conjunctively defined target. However, the visual system cannot avoid the influence of segmentability during the visual search for one feature or conjunction search when one target feature is unknown: the efficiency and speed of search are decreased for nonsegmentable distributions within distractors.

1.10. Data collection

All three articles selected for the defense describe sets of psychophysical experiments. For the present work, we have run twelve laboratory experiments at the Cognitive Research Laboratory (HSE University, Moscow, Russia) and Visual Attention Laboratory (Brigham and Women's Hospital and Harvard Medical School, Boston, USA). Overall, over 200 observers took part in these experiments.

1.11. Public presentations on the topic

The results of the present work have been publicly presented in 5 talks and 6 posters at 8 international and Russian conferences. These included: Annual Vision Sciences Society Meeting (2018, 2019), European Conference on Visual Perception (2018, 2019), Cognitive Science Arena (2020), Actual Problems of Psychological Science (2018), etc. Three colloquium talks have been presented in the HSE Laboratory for Cognitive Research (2019, 2020), Visual Attention Laboratory at Brigham and Women's Hospital (2019).

1.12. Author's contribution

The author was involved in all research described below: discussed the ideas of experiments, created the stimuli, programmed and ran the experiments, analyzed and interpreted the data, presented the results at conferences, and wrote the manuscripts for publications.

2. Independence and parallelism in the visual processing of ensemble statistics

Article selected for the defense: Khvostov & Utochkin (2019)

In Introduction, we reviewed many studies showing that observers can use information about various features of multiple objects to extract different statistical summaries: the mean value of these features, their variance/range, and numerosity (e.g., Ariely, 2001; Burr & Ross, 2008; Chong & Treisman, 2003; Morgan et al., 2008). Mostly, these works have studied the abilities to extract different summaries in isolation which raises a question of their functional relatedness. In the current work (Khvostov & Utochkin, 2019), we divided this issue into two questions: (1) Independence: Are different ensemble statistics computed by a single cognitive system ("the general statistician"), or are they calculated by independent mechanisms? (2) Parallelism: How are several ensemble summaries coordinated in gaining access to conscious perception: can they be calculated in parallel and without a cost of dividing attention between them? Or does the calculating of two summaries lead to mutual interference?

Both these questions have been previously addressed to some degree using two major approaches to studying various domains in perception and cognition (Khvostov & Utochkin, 2019). The question about independence was studied using the individual-difference approach (Huang et al., 2012; Underwood, 1975; Wilmer, 2008): researchers estimate cross-individual correlations between scores in a set of tasks performed by a group of observers. The presence of correlation signifies favor of a common source of noise (variance) implying a common mechanism involved in both tasks. Whereas the absence of correlation means the opposite: different sources of noise for the tasks and different cognitive mechanisms. Using this approach, Yang and colleagues (2018) found the absence of the cross-observer correlation between performance scores of mean and variance calculations (both for size and orientation). They concluded that these two summaries were calculated by independent mechanisms. A similar approach was used in a study by Lee and colleagues (2016) who tested the precision of estimates of mean circle size, numerosity, and total area. Their results showed that mean size and numerosity are independently calculated summaries (whereas total area can be derived from these two parameters).

Parallelism of various mental operations is often investigated using the pre/post-cue paradigm (Khvostov & Utochkin, 2019). For example, a couple of targets (objects, sets, etc.) are shown to observers who should report one of them. In a precue condition, observers are informed in advance which target they should report and can deploy all attentional resources to process that target. By contrast, in the postcue condition, observers have no prior information as to which target they will be asked. They get this information only after the target presentation, so they should divide attention between two targets during the presentation. If performance is better in the precue compared to the postcue condition, it means that two processes interfere and compete for the limited-capacity bottleneck at some point. If there is no such cost in performance, then we can conclude that two processes underlying processing both targets can be done in parallel. The parallelism of mean and numerosity calculations was studied by Utochkin and Vostrikov (2017). In an experiment where observers should calculate two summaries for the same set of circles, they showed no cost in performance between pre/postcue conditions. Therefore, the authors concluded that numerosity and mean can be calculated in parallel. They also probed the independence between these calculations using the individual-difference approach described above. Based on the absence of the cross-observers correlations, Utochkin and Vostrikov (2017) concluded that the mean and numerosity calculations are done by independent mechanisms.

Our study aimed to test and strengthen these conclusions by probing the independence and parallelism of different summary statistics using a new more sophisticated method, the dual task. The previously described pre/postcue paradigm implies that observers indeed try to divide their attention between two summaries in the postcue condition (and do not use some alternative strategies). We developed the dual-task paradigm which allowed us not to rely on this assumption and directly test whether observers can calculate two summary statistics simultaneously by forcing them to report both summaries in each trial (unlike in the postcue condition where they should report only one randomly chosen summary in each trial). The dual-task paradigm also allows us to test independence not only via the indirect cross-observer correlations but also via the direct correlation between two responses in the same trial. If two summaries are calculated by different mechanisms, the accuracy of two responses from the same trial should not correlate with each other. In the opposite case, we expect a positive correlation: if a "general statistician" grasps summaries badly, it should affect both responses in a trial in a similar way and vice versa. One more argument in favor of using this correlation analysis concerns parallelism testing: it can detect a strategy of non-parallel allocation of resources. If observers cannot calculate both summaries simultaneously, they can focus on summary #1 in one trial and on summary #2 in another trial. If it is the case, we should obtain a strong negative correlation between responses in the same trial. Apart from strengthening the conclusions, we also wanted to broaden them. Previously, the mean-numerosity and mean-range relationship were investigated in different studies using different methods and stimuli. Here, we tested both relationships in one study using the same method.

In Experiment 1 (N=24), we tested functional relatedness between the perception of mean size and numerosity. It consisted of several blocks. In a dual-task block, observers were shown a set of circles (from 7 to 36) for 500 ms. After a 200 ms blank screen, they had to report the mean size of the circles and then the number of the circles (or in an opposite order which was counterbalanced). For reporting the mean size, they had to adjust a probe circle using a mouse wheel. To report the numerosity, they had to adjust a probe's numerical value. As a baseline for the performance in the dual-task block, observers participated in two single-task blocks. The only difference between the single and the dual tasks was the requirement to report only one predefined statistic in the former case.

Our primary dependent variable was accuracy, i.e., the normalized absolute error, which we calculated as Error = |Observer's response – Correct response|/Correct response. First, we probed independence via trial-by-trial correlations between two responses within the same dual-task trials for each observer separately. Nineteen out of twenty observers showed no evidence of such correlation. We also ran this analysis for the signed errors (to check whether underestimating numerosity leads to overestimating the mean, as a formula for the average from regular statistics predicts): all twenty observers showed the absence of the correlation. Second, we ran a cross-correlation analysis of average errors in mean size and numerosity reports across observers. As we had three measurements of the average error for each statistic, we calculated three correlations: between the single-task measurements of mean size and numerosity, between first responses in the dual task, and between second responses in the dual task. All three analyses showed no evidence of correlation. Note that auto-correlations of each ensemble summary under different conditions (e.g., mean size in the single task and the dual task) were very high. It shows that it is possible to detect a correlation in our experiment (using our sample size, etc.) when it exists. Also, it provides evidence in favor of the reliability of our measurements. Thus, both macro- (across observers) and micro-level (across trials within each observer) analyses showed that mean size and numerosity are calculated by independent mechanisms. Third, to probe parallelism we compared the accuracy in the dual task with corresponding single-task measurements. This analysis revealed that the error for the first response in the dual task was no different from the error in the single task (both for numerosity and mean size). The accuracy of the second response was worse than the accuracy of the first response and from the single task. It is likely to be explained by memory interference at the recall stage. Therefore, there was no substantial cost of dividing attention between the two summaries which means that mean size and numerosity can be calculated in parallel.

Experiment 2 (N=19) was dedicated to broadening the conclusion of Experiment 1 to another pair of summary statistics: mean size and range/variance. The design and procedure of these two experiments were similar except that the numerosity report was replaced by range. Therefore, each display always contains sixteen circles but the mean size and range of their sizes change from trial to trial. The range adjustment was performed on a set of 16 circles with a fixed mean size. Rotating a mouse wheel increased or decreased the diversity of the circles' set.

We have run the same three analyses to probe the independence and parallelism of mean size and range calculations. As in Experiment 1, the trial-by-trial correlations of two responses from dual-task trials showed the absence of correlation for all observers, both for absolute and signed errors. Cross-correlations between average errors for mean size and range judgments were low and nonsignificant, while the auto-correlations for estimates of the same summary in different conditions were high. These results provide strong evidence in favor of independent calculations of mean size and range. Note that the obtained absence of correlation unequivocally signifies in favor of independent calculations while the presence of correlation could have alternative explanations: either these two tasks are implemented by the same mechanism, or a common mechanism critically affects two different mechanisms implementing these tasks. The parallelism analysis also revealed results similar to those from Experiment 1: the average error of the second response in the dual task was worse than that of the first response or in the single task (probably, due to memory distortion coming from the serial order of recall). While the average error of the first response in the dual task was equal to that from the single task. No cost of dividing attention between mean and range calculations lets us conclude that these two summaries can be calculated in parallel.

Overall, we confirmed the results of the previous work and provided stronger evidence in favor of independent and parallel calculations of several ensemble statistics (Lee et al., 2016; Utochkin & Vostrikov, 2017; Yang et al., 2018). The independence of calculations gives two important insights into ensemble perception. First, different sources of noise in calculations of different ensemble statistics can be taken as evidence that mean size, numerosity, and variance can be calculated by the different cognitive mechanisms which can refer to distinct (or at least partially non-overlapping) neural networks. For example, the independency between the mechanisms for numerosity and other statistical calculations can be illustrated by studies showing that numerical functions, including numerosity estimation, are associated with activity in the parietal cortex (especially the intraparietal sulcus (Dehaene et al., 2003; Nieder & Dehaene, 2009)) and the prefrontal cortex (Nieder & Dehaene, 2009) while the processing of shape statistics (not related to numerosity) is associated with activity in the parahippocampal place area and lateral occipital area (Cant & Xu, 2012). Note that the statement about different neural networks is speculation now and should be verified directly in a separate neurophysiological study. Second, independence also means that the mechanisms doing these calculations are blind to each other and do not use the results of each other work for their calculations. It can be counter-intuitive because regular statistics teach the opposite thing. For example, to compute the mean you need to sum up all the elements and divide them by their number, i.e., computation of the mean uses the results of numerosity calculation by definition (similar things can be said about variance computation). Our study shows that this is not the case for the computation of visual statistics: somehow, the visual system calculates mean size without the knowledge of numerosity.

Parallelism of different summaries computations is consistent with this view: if different summaries are calculated using non-overlapping neural mechanisms, it is easier to do several calculations simultaneously without interference between them. Our results are consistent with the view that, while dealing with many objects, the visual system builds quite a rich summary representation roughly representing the distribution of their features (Chetverikov et al., 2016, 2017a, 2017b, 2017c; Kim & Chong, 2020; Oriet & Hozempa, 2016). This rich representation can be used for the extraction of different pieces of statistical information about the set of objects: mean, numerosity, and variance. How else can this representation be used? The next studies will explore the question of how the visual system uses the information about the distribution of features for visual tasks such as rapid visual categorization and visual search.

3. Ensemble-based segmentation of multiple feature conjunctions

Article selected for the defense: Utochkin, Khvostov, & Stakina (2018)

If the visual system can represent the whole distribution of objects' features (Chetverikov et al., 2016, 2017a, 2017b, 2017c; Kim & Chong, 2020; Oriet & Hozempa, 2016), this distributional representation can be used not only as a basis for explicit estimation of statistical summaries but also for performing everyday cognitive tasks. One idea is that the visual system uses the information about a feature distribution as a cue for rapid visual categorization of groups of objects (Utochkin, 2015; Utochkin & Tiurina, 2014). This process can be explained using the following example. When you watch a soccer match, you can instantly see that there are two visually different groups of players: e.g., one team is in green, and another team is in red. The visual system can easily do this rapid visual categorization because it has access to the color distribution of objects in a field. This distribution has two clear peaks (red and green), each corresponding to players from one team. This type of distribution is called *segmentable*. According to the theory of rapid segmentation and categorization (Utochkin, 2015), this kind of distribution leads to the successful categorization and segmentation of multiple objects into several subsets. Another type of distribution (called *nonsegmentable*) contains one wide peak or is uniform which favors all items being perceived as a single set rather than categorical subsets. If we return to our soccer example, nonsegmentable distribution can be illustrated by many players in differently colored t-shirts: red, yellow, orange, green and their shades. In this case, we cannot see any segmented subsets – it is just one bitty group of players.

The theory of rapid segmentation and categorization was tested in a visual search study where observers searched for an odd-one-out target among distractors that had different feature distributions. When the distractors had a segmentable distribution (e.g., 0°, 22.5°, and 45°), the target (e.g., -45°) was found slower than in the case of a nonsegmentable distribution (0°, 5°, 10°, …, 45°). This result was explained by the fact that in the segmentable case, distractors are grouped into several subsets which observers should inspect serially to reject them as non-targets. In the case of a nonsegmentable distribution, all distractors belonged to one group and could be rejected all at once. The importance of the distribution shape for explicit rapid categorization was also shown using various versions of ensemble tasks requiring summary statistical judgments for subsets (Im et al., 2021).

All the aforementioned studies investigated the role of feature distribution in rapid categorization and segmentation along a single feature dimension. In real-world perception, however, multiple objects rarely show variation, grouping, or segmentation along a single dimension. Often, objects vary along many different dimensions, forming individual feature conjunctions. In the current study, we wanted to test whether the shape of distribution plays an important role in the process of segmentation and categorization of multiple objects defined by a combination of several feature dimensions (i.e., feature conjunctions).

We used a texture discrimination task where observers were presented with an 8×8 array consisting of white lines (64, in total) with different lengths and orientations. We manipulated the distribution of these two dimensions. It could be either segmentable when the distribution consisted of only two extreme feature values (example for orientations: lines tilted only by 11° or 86°), or nonsegmentable when the distribution consisted of two extremes and many intermediate values between them (lines with orientations varied between 11° and 86° with a step of 5°). The 64-cell field was divided in half by an imaginary meridian that could be either horizontal or vertical. Each half contained lines with the same distributions of lengths and orientations as separate dimensions while the conjunctions of these features were distributed differently, providing various length-orientation correlations. In one half of the texture, orientation and length were correlated positively: longer lines were also flatter (and vice versa) – in the other half, they were correlated negatively: longer lines were also steeper. We presented these lines for 200 ms which was followed by a mask (a noisy set of overlapping white lines of different orientations and lengths) for 200 ms. Then observer had to respond whether the boundary between two differently correlated groups of lines was horizontal (upper and lower halves were different) or vertical (left and right halves were different). Therefore, we tested whether observers could rapidly perceive lines with different correlational signs as of different types of objects, i.e., to categorize them.

In Experiment 1 (N=5, experienced observers), we compared performance in two conditions: segmentable (both orientation and length distributions were segmentable) and nonsegmentable (both distributions were nonsegmentable). Also, we manipulated the length-orientation correlation coefficient in the textures – it was ranging from −1.00 to 1.00 with steps of 0.25. It was done by changing the proportion of each feature conjunctions. For example, the correlation approximated by −0.25 was provided by 5/8 of the long lines being steeper and 3/8 of the long lines being flatter. We built the psychometric functions (x-axis – correlation coefficient, y-axis – the proportion of correct responses) for each observer separately for segmentable and nonsegmentable conditions. We fit normal cumulative density functions and analyzed their σ^2 - the variance of the normal distribution which characterized the discriminability of the stimulus. The main result of the experiment was that all five observers showed much smaller σ^2 (better discrimination) in the segmentable condition compared to the nonsegmentable one. It means that the presence of the distinct peaks in the distributions of orientation and length indeed provides better categorization. Observers can clearer see the boundary between two regions with different types of objects.

In Experiment 2 (N=21) we wanted to more closely study this segmentability effect on rapid visual categorization. First, we tested several stimulus presentation durations (from 100 to 500 ms) to reveal what kind of processes are behind this texture discrimination: slow local focused attention (Myczek & Simons, 2008) or fast global distributed attention (Chong & Treisman, 2003). Secondly, we wanted to know how this segmentation effect originates from the more basic segmentability properties of separate feature distributions. Is it enough if only one feature is segmentable or this effect is produced only if both distributions are segmentable? Therefore, instead of two segmentability conditions, we orthogonally manipulated segmentability and had four conditions: "both" (both orientation and length distributions were segmentable), "orientation" (orientation was segmentable, length was nonsegmentable), "length" (length was segmentable, orientation was nonsegmentable), "none" (both distributions was nonsegmentable). In this experiment, we used only two, extremely opposite correlations of 1.0 and -1.0 and calculated the measure of sensitivity to orientation of texture boundary (d' as in the signal detection theory). The results showed that the segmentability effect on the d' occurs quite early (at 200 ms) and does not change during longer presentations. It suggests that the process of slow focused attentional sampling plays a small to nothing role in this effect. This result is more consistent with fast processes of distributed attention which is in line with most views on ensemble statistics (e.g., Chong & Treisman, 2005a). The second important result is that only the condition where both length and orientation distributions were segmentable provided rather good performance (d' around 0.8-0.9), whereas performance in the rest of the conditions was much poorer (d' below 0.3) suggesting that the segmentability of only one dimension is not enough to provide a good basis for categorization of multiple conjunctions. We will address this result in Experiment 4.

Experiment 3 (N=23) was dedicated to more direct testing whether the segmentability effect reflects the work of global ensemble processes or local focused attentional processes. Instead of analyzing all presented objects, there is a possibility that observers do the task using a strategy based solely on focused attention (though it contradicts the results of Experiment 2 with segmentability effect raising at 200 ms). They could just compare two near elements from different halves along one of the boundaries (e.g., horizontal) – if they are "in agreement" (e.g., two long vertical lines or one long vertical and one short horizontal lines), an observer should respond that the boundary is opposite (i.e., vertical), if they are not in agreement – the boundary is horizontal. To test it, we compared the performance in the full-texture condition (the same as in the previous experiment) and near-boundary condition (where we presented only lines along horizontal and vertical meridians of the array). If observers use focused attention to perform the task, their performance should not decrease under the near-boundary condition since they should use only two elements near one of the boundaries. But if the global processes are involved, the performance in the full-texture condition should be higher since we provide more statistical information. The results showed the advantage of full-texture over the near-boundary condition in all segmentability conditions providing another argument in favor of explaining the segmentability effect by the global process of distributed attention rather than the local process of focused attention.

If focused attention does not play a significant role in the rapid categorization of multiple conjunctions, how do observers perform the task? We think that observers use global mechanisms of distributed attention to implement a strategy that we call half-split. We suppose that observers try to select all objects from one half of a first feature distribution and then compare the groups of selected objects along a second feature dimension to find any average difference. For example, an observer can select a subset of long lines and check where there is a difference between average orientations. This global ensemble strategy might be quite fast and give the impression of differences between patches without computing the full correlations. Feature segmentability can facilitate this process in two ways. First, by simplifying the selection process: it is much easier to select all long lines if they are the same length, and all not-to-be-selected lines are short. Second, this is accomplished by increasing the average difference and decreasing variability within the groups along the second feature dimension. In the segmentable case, the average difference in orientation between two patches is 75° (the variance is 0 because all lines are of the same orientation), in the nonsegmentable case, the mean difference is 40° (the variance is 12°). These facts can be used to explain the result from Experiment 2 that the segmentability of only one feature is not enough to substantially increase the performance in a rapid categorization task.

Experiment 4 (N=16) was dedicated to testing the half-split hypothesis. We artificially simulated the perfect half-split selection and presented half-textures from Experiment 2 removing from original stimuli half of the items with features drawn from one half of either orientation or a length distribution. Texture discrimination then turns into patch comparison along a single dimension. Will the performance be better without the need for a half-split and selection along the first dimension? And will the segmentability of the second dimension improve performance in this case? The results showed that performance was much better than in all previous experiments (d' = 1.12 in Experiment 4 vs 0.345 in Experiment 2, at 200 ms presentation duration). Also, we found the segmentability effect for the second dimension which was the following: if the half-split was made based on length (only longer lines were presented), the segmentability of orientation improved the performance drastically (and vice versa). Therefore, this experiment showed support for the half-split hypothesis. Two theoretical difficulties predicted by this strategy were empirically tested. First, observers indeed have problems with the correct selection of objects from one half of the first dimension (when this selection was made artificially or when this dimension is segmentable, performance becomes much better). Second, comparison of average values along the second dimension can also be a problem, since segmentability of this dimension increases the performance.

Overall, four experiments showed that the visual system can use the shape of the distribution for performing rapid visual categorization of multiple objects. It can be done even when groups of objects are defined as conjunctions of several features. Observers likely use the half-split strategy to select a group of objects from one half of a first dimension and compare these selected objects along a second dimension. Such distributional property as segmentability (the presence of several peaks with a large gap between them) makes this hard process possible because it facilitates the selection of objects along the first dimension and increases the average difference (and decreases within-group variance) along the second dimension.

4. The role of the distributional shape in conjunction visual search

Article selected for the defense: Utochkin, Khvostov, & Wolfe (2020).

This study is a logical continuation of the previous research. Here, we aimed to test whether the distributional segmentability of two feature dimensions works similarly for another important everyday cognitive task – visual search. Like texture segmentation, visual search requires processing multiple objects at the same time to respond whether a target (an object with predefined features or the odd one out) is present among distractors.

In Experiment 1 (N=12), we presented observers with many lines with different orientations (from very shallow to very steep) and colors (from red to blue). Their task was to find a predefined conjunction target (e.g., red shallow line) among distractors (e.g., red steep and blue shallow lines) as fast as possible. As in the previous study, our main manipulation concerned the shape of the distribution. It could be either nonsegmentable (orientation ranging between 10° and 80° with a step of 10° and colors from the CIE Lab color wheel from 270° (blue) to 360° (red) with a step of 12-13°) or segmentable (orientation: only 10° and 80°, colors: only 270° and 360°). We manipulated segmentability orthogonally and had four conditions: "both" (both features are segmentable), "orientation" (orientation was segmentable, color was nonsegmentable), "color" (orientation was nonsegmentable, color was segmentable), and "none" (both features was nonsegmentable). As in any classical visual search study, we also varied the number of objects (i.e., set size: 9 vs. 17) and the presence of a target on the screen (present vs. absent). We plotted a set size × RT function, its slope tells us how efficient the deployment of attention was (how many ms observer spends to process each additional item). Slopes can differ from around 0 values (meaning the search was close to parallel: attention goes straight to a target location regarding the number of distractors) to very high values (meaning that the search was completely serial: observers should inspect item by item sequentially unless he/she find the target). The results revealed the absence of the segmentability effect. In all segmentability conditions, observers showed quite an efficient search: slopes were 9-14 ms/item which is close to parallel.

Experiment 2 (N=15) aimed to replicate this result using another pair of features from the previous study: length and orientation. We used the same experimental design as in Experiment 1 and obtained similar results: no effect of feature segmentability and very efficient search in all conditions (slopes: 0-5 ms/item).

Given our previous finding of a considerable segmentability effect on texture discrimination (Study 2), this result looks surprising. But the lack of segmentability effect on conjunction search can be explained using the Guided Search model of visual search (Wolfe, 1994, 2021). This theory proposes that the visual system has access to various retinotopic feature maps (for color, orientation, size, etc.). Each place on such a feature map has activation proportional to the similarity of the feature of an object in this place to the target feature. Summing the activations from many feature maps, the visual system builds the attention priority map which is used for guidance of attentional deployment during visual search – the selective attention first "goes" to a place with the highest activation. Thus, the Guided search model can explain the efficiency of our searches from the first two experiments by saying that attention is guided simultaneously toward red and toward steep in a search for a red steep item. As all elements in a display are either steep or red, only the target is both steep and red. This double dose of guidance attracts attention to the target location in a similar fashion both in the segmentable and nonsegmentable conditions. In the latter case, the target gets a double dose of guidance, while distractors get less because as one guiding feature (e.g., steepness) gets stronger, the other (redness) gets weaker (due to opposite directions of correlation between color and orientation for target and distractors). This model can explain why stimuli that create the strong effect of segmentability in the case of texture segmentation do not do the same for conjunction search. Also, this model predicts that segmentability will play an important role in conjunction search when observers do not know all the target features.

In Experiment 3 (N=13), observers performed the so-called subset search task in displays like in Experiment 1. The only difference was that there was no predefined target so observers searched for an odd-one-out orientation in a subset of items with a known color (e.g., a unique orientation within a red subset). It means that the target (as well as the distribution of distractors) was changed from trial to trial: in one half of trials, a target could be a red steep line (distractors were red shallow and blue steep lines), in the other half, a target could be a red shallow line (distractors were red steep and blue shallow lines). As in previous experiments, we had four segmentability conditions (“both”, “orientation”, “color”, “none”) and one more condition for comparison, standard conjunction search (“both” condition from Experiment 1). The results again showed that searches in all conditions were quite efficient (slopes are 5-14 ms/item). But unlike previous experiments, the segmentability had a considerable effect on average RT: observers were much slower (~200-300 ms) in those conditions where one or both feature distributions were nonsegmentable (“color”, “orientation”, and “none”). Note, that this effect mirrored the effect from Study 2 (Experiment 2) where the only condition with both segmentable features produced good performance while all others were not. Our explanation of this effect is that observers spend the additional time to figure out what the target orientation in this probe is: it is much harder to do if at least one feature is nonsegmentable. If color is nonsegmentable, it is harder to select the color group to search for the odd-one-out orientation among them. If orientation is nonsegmentable, the odd-one-out element is more similar to distractors from a color subset. But once an observer figures out the target orientation, the task becomes a typical conjunction search, where observers can deploy attention with similar efficiency in segmentable and non-segmentable conditions. Hence, we observed no change in the slopes of the RT x set size functions, but the significant effect of segmentability on the average RTs. Therefore, we could detect the effect of segmentability in a search task where the role of preliminary categorization is increased.

Experiment 4 (N=14) tested another prediction from the Guided search model regarding segmentability. The absence of the segmentability effect in Experiments 1-2 was explained by the fact that the Attention priority map sums activation patterns of two feature maps and got a similar result for segmentable and nonsegmentable cases. But this similar pattern comes from very different activation on Feature maps. In the segmentable conditions, all places on feature maps are either highly activated (objects have the same feature as the target) or weakly activated (objects have completely different from the target feature value). In contrast, in the nonsegmentable conditions, both feature maps show a gradient of activations from high to low. Thus, the Guided search model can predict a rather counterintuitive result: feature search in the nonsegmentable condition (when the target is defined either by a unique color or by a unique orientation only) can be slower and less efficient than conjunction search for the same stimuli. This prediction provides a good test for our explanations because it goes against the classical result by Treisman and Gelade (1980) that feature search is faster than corresponding conjunction search. We had slightly different stimuli but a similar procedure and experimental design to Experiment 1. Out of four segmentability conditions from conjunction searches of Experiment 1, we took only "both" and "none" and compared them with two component feature searches (orientation and color) both performed on non-segmentable stimuli. In all conditions, observers showed a search for the same target (e.g., a white vertical line). In orientation search, observers had to search among differently oriented white lines (nonsegmentable distribution): in color search, the observers had to search among differently colored vertical lines (from white to red, nonsegmentable distribution); in segmentable conjunction search, they had to search among red vertical and white horizontal lines (segmentable distribution); in nonsegmentable conjunction search, they had to search among differently oriented and differently colored lines (nonsegmentable distribution). The results showed that both segmentable and nonsegmentable conjunction searches were much faster and more efficient than both feature searches (especially, orientation) which are strong empirical evidence in favor of our explanation of the whole set of experiments.

Experiment 5 (N=26) was dedicated to replicating the main results of this series of experiments with a larger sample of observers (to show that the absence of segmentability effect was not due to lack of power) and broaden the conclusions to denser, texture-like displays with a larger number of objects. We used the same experimental design as in Experiment 1 except that we added two additional set size conditions (4 set sizes in total: 9, 17, 33, 65). The results of this experiment strongly replicated our results from the previous experiments. When observers know both target features, all segmentability conditions provided very efficient searches (7-10 ms/item) without any big difference between conditions.

Overall, this set of experiments showed how the visual system uses the information about the shape of the feature distributions to perform a visual search for a target defined by two features. When both target features are known by observers, the visual system can bypass the absence of clear categorical groups among distractors and efficiently find a target using parallel guidance by the sum of two feature maps. But if observers do not know at least one of the target features, the absence of segmentability decrease the speed of the search significantly. In other situations where the visual system cannot use guidance by several features (e.g., in feature searches) segmentability of feature distribution was shown to be a very important prerequisite for fast efficient search.

5. Conclusion

The main question of this dissertation was how the visual system can use the rich ensemble representation of feature distributions. As a result of our research, we came to the following conclusions.

First, the visual system can effectively extract different statistical summaries out of this rich representation. We obtained solid evidence that it can be done in parallel, i.e., without any interference between calculations of two different summaries. Consistently with this, different ensemble summaries are read out from the general representation by independent cognitive (and likely, neural) mechanisms which means that calculations of each ensemble summary are done independently from the others.

Second, the shape of the feature distribution can be used as a cue for rapid visual categorization and segmentation, even in such a complicated case when different subsets are defined by multiple conjunctions of several features. Observers likely use the half-split strategy: they select a group of objects from one half of the first dimension and compare these selected objects along the second dimension. The segmentability of both feature distributions is a prerequisite for successful performance in such a task because it facilitates the selection of objects along the first dimension and increases the average contrast along the second dimension. This segmentability effect reflects the work of a globally distributed attentional process as it occurs rather early and the performance benefits from the presence of full-texture stimuli compared to only local elements near the boundary.

Third, the conjunction search task turns out to be in a more complicated relationship with segmentability. When observers know both target features, the visual system can bypass the absence of distributional segmentability among distractors and efficiently find a target using parallel guidance by the sum of two feature maps. But if at least one feature is not known, the nonsegmentability of the distribution drastically decreases the search speed. In other situations where the visual system cannot use guidance by several features (e.g., in feature searches) segmentability of feature distribution was shown to be a very important prerequisite for fast efficient search.

It should be noted that the conclusions about the usage of the rich ensemble representation of feature distributions can be generalized to many different feature dimensions. We did not blindly assume that ensemble perception works similarly for all features but tested it on size, color, and orientation. Moreover, each study had at least one experiment with size-orientation pair of features (only size in Study 1) so the different results between studies cannot be explained by the choice of different features. Therefore, our conclusions can be considered as solid, replicable, and generalizable on a range of visual features.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychological Science*, *19*(4), 392–398. https://doi.org/10.1111/j.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*, *106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, *12*(2), 157–162.

Bauer, B. (2009). Does Stevens's Power Law for Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record*, *59*(2), 171–185. https://doi.org/10.1007/BF03395657

Burr, D., & Ross, J. (2008). A Visual Sense of Number. *Current Biology*, *18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

Cant, J. S., & Xu, Y. (2012). Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex. *Journal of Neuroscience*, *32*(22), 7685–7700. https://doi.org/10.1523/JNEUROSCI.3325-11.2012

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017a). Chapter 5 - Learning features in a complex and changing environment: A distribution-based framework for visual attention and vision in general. In C. J. Howard (Ed.), *Progress in Brain Research* (Vol. 236, pp. 97–120). Elsevier. https://doi.org/10.1016/bs.pbr.2017.07.001

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017b). Representing Color Ensembles. *Psychological Science*, *28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017c). Set size manipulations reveal the boundary conditions of perceptual ensemble learning. *Vision Research*, *140*, 144–156. https://doi.org/10.1016/j.visres.2017.08.003

Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate): Distributed versus focused attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 634–638. https://doi.org/10.1002/wcs.136

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 12.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

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*(2), 289–301. https://doi.org/10.1016/j.actpsy.2011.08.002

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition*, *20*(2), 211–231. https://doi.org/10.1080/13506285.2012.657261

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/S0042-6989(97)00133-8

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). THREE PARIETAL CIRCUITS FOR NUMBER PROCESSING. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep32210

Gardelle, V. de, & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13341–13346.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science*, *17*(7), 572–576. https://doi.org/10.1111/j.1467-9280.2006.01746.x

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 414–428. https://doi.org/10.1037/a0026314

Im, H. Y., Tiurina, N. A., & Utochkin, I. S. (2021). An explicit investigation of the roles that feature distributions play in rapid visual categorization. *Attention, Perception, & Psychophysics*, *83*(3), 1050–1069. https://doi.org/10.3758/s13414-020-02046-7

Khvostov V., Lukashevich A., Utochkin I. S. (2021). Spatially intermixed objects of different categories are parsed automatically. *Scientific Reports,* 11, 377, 1-8. https://doi.org/10.1038/s41598-020-79828-4.

Khvostov, V. A., & Utochkin, I. S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision*, *19*(9), 3. https://doi.org/10.1167/19.9.3

Kim, M., & Chong, S. C. (2020). The visual system does not compute a single mean but summarizes a distribution. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(9), 1013–1028. https://doi.org/10.1037/xhp0000804

Lee, H., Baek, J., & Chong, S. C. (2016). Perceived magnitude of visual displays: Area, numerosity, and mean size. *Journal of Vision*, *16*(3), 12. https://doi.org/10.1167/16.3.12

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms13186

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4), 6. https://doi.org/10.1167/15.4.6

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A “dipper” function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11), 9–9. https://doi.org/10.1167/8.11.9

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. https://doi.org/10.3758/PP.70.5.772

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, *15*(4), 3. https://doi.org/10.1167/15.4.3

Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision*, *16*(3), 3–3. https://doi.org/10.1167/16.3.3

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. https://doi.org/10.1038/89532

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197. https://doi.org/10.1163/156856888x00122

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 13–13. https://doi.org/10.1167/11.12.13

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*(2), 128–134. https://doi.org/10.1037/h0076759

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, *15*(4), 8. https://doi.org/10.1167/15.4.8

Utochkin, I. S., Khvostov, V. A., & Stakina, Y. M. (2018). Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions. *Cognition*, *179*, 178–191. https://doi.org/10.1016/j.cognition.2018.06.016

Utochkin, I. S., Khvostov, V. A., & Wolfe, J. M. (2020). Categorical grouping is not required for guided conjunction search. *Journal of Vision*, *20*(8), 30. https://doi.org/10.1167/jov.20.8.30

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, *146*, 7–18. https://doi.org/10.1016/j.actpsy.2013.11.012

Utochkin, I. S., & Vostrikov, K. O. (2017). The numerosity and mean size of multiple objects are perceived independently and in parallel. *PLOS ONE*, *12*(9), e0185452. https://doi.org/10.1371/journal.pone.0185452

Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by “segmentability”. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 995–1007. https://doi.org/10.1037/xhp0000203

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-I

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, *18*(3), 484–489. https://doi.org/10.3758/s13423-011-0071-3

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, *69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Wilmer, J. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, *21*(6), 561–579. https://doi.org/10.1163/156856808786451408

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is There a Common Summary Statistical Process for Representing the Mean and Variance? A Study Using Illustrations of Familiar Items. *I-Perception*, *9*(1), 204166951774729. https://doi.org/10.1177/2041669517747297

Ying, H., & Xu, H. (2017). Adaptation reveals that facial expression averaging occurs during rapid serial presentation. *Journal of Vision*, *17*(1), 15–15. https://doi.org/10.1167/17.1.15