NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Igor S. Utochkin, Vladislav A. Khvostov,
Yulia M. Stakina

ENSEMBLE-BASED SEGMENTATION IN THE PERCEPTION OF MULTIPLE FEATURE CONJUNCTIONS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY
WP BRP 78/PSY/2017

SERIES: PSYCHOLOGY

Igor S. Utochkin[1]*, Vladislav A. Khvostov*[2]*, Yulia M. Stakina*[3]

ENSEMBLE-BASED SEGMENTATION IN THE PERCEPTION OF MULTIPLE FEATURE CONJUNCTIONS[4]

Although objects around us vary in a number of continuous dimensions (color, size, orientation, etc.), we tend to perceive the objects using more discrete, categorical descriptions. For example, in the variety of colors and shapes on a bush, we can see a set of berries and a set of leaves. Previously, we described how the visual system transforms the continuous statistics of simple features into categorical classes using the shape of distribution. In brief, "sharp" distributions with extreme values and a big gap between them are perceived as "segmentable" and as consisting of categorically different objects, while "smooth" distributions with both extreme and moderate features are perceived as "non-segmentable" and consisting of categorically identical objects. Here, we tested this mechanism of segmentation for more complex conjunctions of features. Using a texture discrimination task with texture difference defined as length-orientation correlation, we manipulated the segmentability of length and orientation. We found that observers are better at discriminating between the textures when both dimensions are segmentable. We assume that the segmentability of both dimensions leads to rapid (within 100-200 ms, as our data show) segmentation of conjunction classes which facilitates the comparison between the textures containing these classes.

JEL Classification: Z

Keywords: ensemble summary statistics, segmentation, categorization, texture discrimination, features and conjunctions, sampling.

---

[1] "Laboratory for cognitive research", National Research University Higher School of Economics, Russian Federation. E-mail: iutochkin@hse.ru

[2] "Laboratory for cognitive research", National Research University Higher School of Economics, Russian Federation. E-mail: vkhvostov@hse.ru

[3] "Laboratory for cognitive research", National Research University Higher School of Economics, Russian Federation. E-mail: stakina@hse.ru

[4] The study was conducted within the Basic Research Program of the National Research University Higher School of Economics in 2017 (TZ-59).

Introduction

Our capacity to attend to objects and store them in the working memory for deep processing is very limited (Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988; Scholl, 2001). However, in everyday perception we often encounter hundreds of objects at one time but do not have difficulties in seeing them all. How can these hundreds of objects survive the severe limits of the processing bottleneck? One possible answer is that the visual system can represent multiple objects in the compressed form of ensemble summary statistics (Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016). Instead of the precise encoding of each individual item, a summary of many objects can be computed to describe them together. This gives a number of advantages. First, it provides more accessible and perhaps more precise information about the objects than any lossy individual representation (Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Parkes, et al., 2001). Secondly, representing statistical summaries of many objects requires as much attention as representing individual properties of a single object (Huang, 2015) and can be performed without focusing on each object (Alvarez & Oliva, 2008; Chong & Treisman, 2005a; Huang, 2015; Robitaille & Harris, 2011; Utochkin & Tiurina, 2014; but see Allik et al., 2013; Myczek & Simons, 2008; Maule & Franklin, 2015). Thirdly, ensemble statistics probably do not suffer (Alvarez & Oliva, 2008; Epstein & Emmanouil, 2017) or even benefit (Bauer, 2017) from a concurrent task requiring individual objects to be attended to and maintained in working memory (but see Jackson-Nielsen, Cohen, & Pitts, 2017). Ensemble summaries can be computed for basic sensory domains: size (Ariely, 2001; Chong & Treisman, 2003), orientation (Alvarez & Oliva, 2008; Dakin & Watt, 1997), color (Huang, 2015; Maule & Franklin, 2015; De Gardelle & Summerfield, 2011), brightness (Bauer, 2009), direction and speed of motion (Emmanouil & Treisman, 2008; Watamaniuk & Duchon, 1992) and for complex perceptual features (Haberman & Whitney, 2007; Yamanashi-Leib & Whitney, 2016). Different types of descriptives, such as the mean, variance and numerosity are available as a result direct perceptual processing, as they are prone to adaptation aftereffects (Burr & Ross, 2008; Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Norman, Heywood, & Kentridge, 2015).

One relevant point that is often mentioned in the literature is the link between statistical representations in vision and perceptual organization. It has been shown in previous studies that the statistical properties of an image predict how the elements of that image are grouped or segregated. For example, feature variance affects the precision of visual averaging, which can be interpreted as the ease with which a holistic representation of a group of objects is made (Im & Halberda, 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014). On the other hand, the discriminability of two different groups depends on their mean difference and variances (Corbett et al., 2012; Fouriezos, Rubenfeld, & Capstick, 2008). Finally, results of visual search experiments show that statistical outliers – elements whose difference from the mean is substantially greater than the overall variance – are easily segregated from the surroundings, drawing attention and saccades (Avraham, Yeshurun, & Lindenbaum, 2008; Haberman & Whitney, 2012; Nothdurft, 1992, 1993; Palmer, Verghese, & Pavel, 2000; Rosenholtz, 2001; Rosenholtz, Huang, & Ehinger, 2012a; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012b). It has also been shown that the visual system tends to downplay or exclude outlying values when computing the mean, thus protecting organization by similarity (Haberman & Whitney, 2010).

Segmentability: Distribution-based segmentation and categorization

Our recent work (Utochkin, 2015; Utochkin & Yurevich, 2016) focused on testing the idea that a set of multiple items can be grouped or segregated based on the shape of the feature distribution. Within a fixed range of feature variation, items can be rapidly categorized as same-type objects or as different-type objects, depending on the peaks and gaps in the internal distribution. This theory explains how a distribution of *continuous* visual features can be rapidly transformed into *discrete* classes of objects in perception. If one peak is present, then the set is more likely to be perceived as consisting of same-type objects. By contrast, if several peaks are presented and interleaved with gaps then the items represented by each peak are more likely to form different-type objects (Utochkin, 2015). Correspondingly, same-type objects are better grouped, while different-type objects are readily segmented, even if the types are intermixed in the space. For example, leaves on a tree can widely vary in the fall from green to red as extremes, but individual shades can be intermediate. The presence of these intermediate shades makes the transition between green and red smooth, so this produces a single-peak distribution recognized as a set of one-type objects. By contrast, in summer, leaves and ripe berries also vary between green and red. But the transition between the extremes is much more abrupt, so one would more easily see this set as two overlapping sets of different-type objects. Utochkin and Yurevich (2016) tested this theory in three visual search experiments. Their participants had to search a size (Experiment 1) or orientation (Experiments 2 and 3) singleton among homogeneous or heterogeneous distractors with different sizes or orientations, respectively. While the range of feature variation and set sizes were carefully controlled, the transition between extreme feature values was manipulated. Transition could be *sharp*, assuming that feature values were distributed with a relatively big step (e.g., 0º, 22.5º, and 45º of orientation) or extremely (e.g., 0º and 45º). Transition could also be *smooth*, assuming a much smaller step between neighboring values (e.g., 0º, 5º, 10º, …, 45º). Utochkin and Yurevich (2016) found that, despite being the most heterogeneous condition (Duncan & Humphreys, 1989), smooth distractor distributions yielded faster search than any of the sharp distributions. Among the sharp conditions, the one with three values was more difficult for the search than the one with only two extremes. Therefore, search efficiency was related to heterogeneity non-monotonically. Utochkin and Yurevich (2016) suggested that the size of the transition was critical and introduced a concept of *"segmentability"* to explain the non-monotonic effect of smoothness. The set with sharp transitions between the features provides internal distributions with peaks corresponding to each presented value, and large gaps between them. That should lead to segmentation of the set into categorically different subsets (Utochkin, 2015), so that each subset is analyzed as a separate chunk and rejected serially, making search slower (Duncan & Humphreys, 1989; Humphreys & Müller, 1993;

Müller, Humphreys, & Donnelly, 1994). When a transition is smooth, there are no large gaps in the internal distribution, so the visual system collects all local featural peaks under a single-peak global distribution. Although this distribution has large variance, it is more likely to be represented as one group and rejected at once in visual search.

Segmentation of multiple conjunctions: A general framework

In real-world perception, multiple objects rarely show variation, grouping, or segmentation along a single dimension. Each object varies in many features, forming an individual feature conjunction; taken together, multiple objects can provide a huge variety of conjunctions (Tsotsos, 1987). The variety of conjunctions as a function of their constituent feature statistics can be described in terms of inter-feature *correlation*. The correlation (or any other concordance measure) is a good way to estimate how likely certain features in one dimension go with certain features in another dimension. An analogy can be made between descriptive ensemble statistics (e.g. average size) in global feature discrimination, and correlational statistics in global conjunction discrimination. Comparing two sets of dots with different average sizes, we can say where bigger dots prevail. Likewise, comparing two sets with different color-size correlations, we can say where reddish and bigger items prevail. Therefore, our discrimination between conjunction-defined sets of multiple objects is the matter of “seeing” feature correlations.

When working with multiple conjunctions provided by several variable dimensions, can the visual system rapidly segment the classes of conjunctions? Sometimes, even when several separate dimensions are variable, conjunction segmentation is not harder than segmentation along a single dimension. This can happen when the features are highly correlated, which often occurs in the real world. For example, ripe berries on a bush are well distinguished from surrounding leaves not only because they are red, but also because they significantly differ in shape and size. Due to their high feature correlation and segmentability, the berries and the leaves must be perceived reliably as two object populations, because any of the feature distributions can be used for segmentation and no single distribution contradicts the rest. The segmentation task can be reduced to one-feature segmentation and based on a single statistical summary. But the case becomes more interesting when feature distributions overlap between the populations. Imagine that not all berries on the bush are ripe enough, so their color ranges from green to red through shades. Then the color distribution is no longer segmentable. Moreover, the greenish part of the berry color distribution is now shared with the leaf color distribution. Therefore, the overall color distribution is less useful to rapidly segment berries from leaves. Given that shapes are still segmentable enough, will segmentation be efficient? Shape segmentability supports categorical differences between leaves and berries, but the non-segmentable color distribution supports one category, thus potentially interfering with segmentation. Another interesting case is an attempt to segment a relevant item population from stuff where different populations have very similar feature distributions, but these features are correlated differently. Imagine again, that you have two bookshelves filled with reddish and bluish books of A5 to A4 sizes. The proportion of colors and sizes are about equal between the shelves, but shelf 1 has more reddish A4 books and shelf 2 has more bluish A4 books, so color-size correlations are different. If you are looking for a particular book, which is "big red", can you determine with one brief glance which shelf is more likely to have this book? Can you also delimitate the subsets of potential target books from irrelevant ones? How can color segmentability help you in perceiving the proportions of the "big red" books and in subset delimitation? These are the questions that we are going to answer in our study.

The question of how the visual system treats multiple feature conjunctions has rich theoretical links. Perhaps, the most fundamental topic this question is related to the "binding problem" (Cave & Wolfe, 1999; Treisman, 1999). Perceived objects and scenes are presumably represented as elementary features and parts in the early sensory analysis performed by isolated and independent sensory modules (Treisman, 2006; Yantis, 2014) and should be somehow integrated correctly. As complete and momentary binding seems to be very computationally demanding (Tsotsos, 1987), it are prone to some limitations. A theoretical debate concerns the locus of binding limits and strategies that the visual system uses to deal with them (e.g., Di Lollo, 2012; Duncan & Humphreys, 1989; Cave & Wolfe, 1999; Rosenholtz, Huang, & Ehinger, 2012a; Treisman, 2006; Treisman & Gelade, 1980, etc.). The perception of correlated features in multiple items is a question of binding to some degree. How accurately can colors and sizes be ascribed together to the books from our example, given that even separate features are represented quite approximately – as an ensemble summary rather than a set of precise values for each book? Can these correlation statistics be computed very efficiently across multiple regions at one time (Rosenholtz et al., 2012b)? Or does it require a slower accumulation of information across more local samples (Allik et al., 2013; Simons & Myczek, 2008), perhaps because the serial deployment of attention is required for correct binding (Treisman, 2006; Treisman & Gelade, 1980)?

Our study

In this study, we investigate whether ensemble-based segmentation takes place in sets of multiple conjunctions, where the specific visual properties in those sets are defined by a correlation of features. Previous research using ensemble tasks, visual search, and texture discrimination has found that such multi-element displays (sets or textures) can be discriminated to some degree. However, such discrimination is usually worse than in displays where the specific properties are defined by differences in a single feature (Emmanouil & Treisman, 2008; Found, 1998; Rosenholtz et al., 2012b; Treisman, 2006; Treisman & Gelade, 1980; Wolfe, 1992). While some theories explain this by the quite poor discriminability of image statistics of correlated conjunctions compared to differently distributed features (e.g., Rosenholtz et al., 2012b), others suggest a different strategy which is particularly relevant to segmentation issues. Observers can first select a class of items defined by a specific feature and ignore other classes in the same dimension. Then, within the selected class only, they can search for differences in features (Friedman-Hill & Wolfe, 1995; Nakayama & Silverman, 1986; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989) or feature statistics (Chong & Treisman, 2005b) along a different dimension (Chong & Treisman, 2005b). This sort of processing is often labeled “guided search” (Wolfe, 1994) in the visual search literature. But for the implementation of this serial algorithm, the classes should be defined at least along a dimension that had been selected first. Segmentability can therefore influence selection.

What prediction can we make to test whether segmentation of multiple conjunctions occurs rapidly? According to our segmentability theory (Utochkin, 2015; Utochkin & Yurevich, 2016), segmentable sets facilitate the discrimination between sets with different feature correlations, because their statistics make it easier to select local classes for comparison. We ran two experiments to investigate the effect segmentability has on the discrimination of sets (textures) of two correlated features – the lengths and orientations of lines. In Experiment 1, we measured how well observers discriminated between two textures as a function of their segmentability and the difference between their correlation statistics. As in Utochkin and Yurevich (2016), we manipulated segmentability using sharp and smooth transitions of physical feature distributions within a fixed range. In Experiment 2, we manipulated segmentability orthogonally across both feature dimensions. By doing this, we wanted to see whether segmentation along one dimension provides better segmentation in general, even if the second dimension is non-segmentable. We also varied the exposure time in Experiment 2 in order to track the time course of extracting correlation statistics and segmentation.

Experiment 1

Methods

Participants. Five expert observers, including the authors of this article, participated in the experiment. Their age varied between 20 years and 43 years old, median age was 21 years old. All had normal or corrected-to-normal vision and no neurological problems.

Apparatus and stimuli. Stimulation was developed and presented through PsychoPy for Linux (Pierce, 2007). Stimuli were presented on a standard VGA monitor with a refresh frequency of 75 Hz and a 800 × 600-pixel spatial resolution. A 26° × 26° square at the center of the screen was used as the “working” field for presenting stimuli; the rest screen space remained gray. The working field was divided into 8 × 8 = 64 cells by an imaginary grid (each cell side was 3.25°). Each cell was used for positioning a single line element of the display. Within the cell, a line could be randomly jittered within ±0.23° in both horizontal and vertical directions.

We used white lines as items in the sample set. The width of the lines were 0.17°. The length varied between 0.91° and 3.08°. The step of length increment was 0.14°. The orientation of lines varied between 11° and 86° with steps of 5° (the angles were counted from the vertical axis, so the greater angle value corresponds to a flatter orientation). Only the extreme values of both length and orientations were drawn from this range in the “sharp” condition of the experiment (Figure 1A). In the “smooth” conditions all the steps of orientation and length were shown (16 values in each dimension, Figure 1B). That is, sharp displays aimed to model segmentable sets, while smooth ones were intended to model non-segmentable sets (Utochkin & Yurevich, 2016).

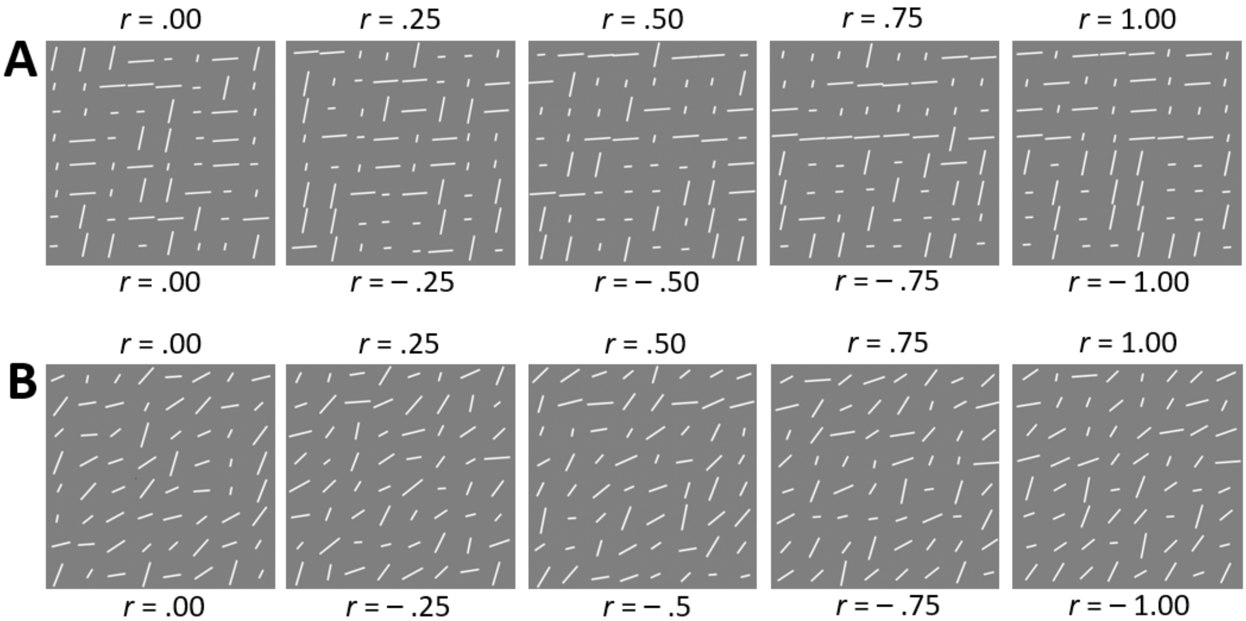


Fig. 1. Example displays with (A) sharp and (B) smooth distributions of features and different correlation coefficients. The examples show positive correlations in the top halves of the

displays and negative correlations in the bottom halves, but inverse stimuli were also present during the experiment.

The 64-cell working field was divided in half (top and bottom, $8 \times 4 = 32$ cells each) by an imaginary horizontal meridian. Each half contained lines with exactly the same distributions of lengths and orientations as separate dimensions. In "smooth" displays, each half contained a double set of lines with 16 values of length and 16 values of orientation; in "sharp" displays, two extreme values of length (the shortest and the longest) and orientation (the steepest and the flattest) were presented 16 times in each half (Figure 1). While the distributions of lengths and orientations were equated between the halves, the conjunctions of these features were distributed differently, providing various levels of length-orientation correlations. These levels can be approximated by correlation coefficients ranging from -1.00 to 1.00 with steps of 0.25. For example, the correlation coefficient of 1.0 means that a longer line is exactly combined with a greater deviation from vertical, and the shorter lines were closer to vertical (steeper). As another example, correlation approximated by -0.25 was provided by 5/8 of the longer lines being steeper and 3/8 of the longer lines being flatter. Lines generated this way were randomly assigned to positions within each half (Figure 1). Within each compared half, the distributions of line-orientation conjunctions were counterbalanced between their inner halves (or quarters of the entire field). This was done to avoid a situation where features from different sides of the distributions are present unevenly between the quarters, thus causing conflicting perceptual grouping.

Both halves of the display in one trial always had the same absolute value of length-orientation correlation, but signs were always opposite (except for the correlation of 0). Figure 1 demonstrates examples of displays with various levels of differences in correlation between the top and bottom halves.

Procedure. Participants were seated approximately 50 cm from the screen. Two of them were instructed to answer which half of a display contains more longer and flatter lines. This means that they were instructed to search for a positive correlation. The rest were instructed to answer which half of a display contains more longer and steeper lines. This means that they were instructed to search for a negative correlation. This was made for counterbalancing possible instruction-associated biases.

Each trial started with the presentation of a fixation point for 500 ms (Figure 2). The sample display was shown during for 200 ms and was followed by a mask (a noisy set of overlapping white lines of different orientations, Figure 2), which was also presented for 200 ms. After that participants had to report which half of the display had contained the relevant set, using “up” and “down” arrow keys on a standard computer keyboard. Responses were followed by feedback informing the observers whether answer had been correct or not. A next trial started by pressing the SPACE bar, so participants could progress at a comfortable pace and take a rest whenever they wanted.

At the beginning of the experiment, the participants did a training session consisting of eighteen trials for familiarization with the task.

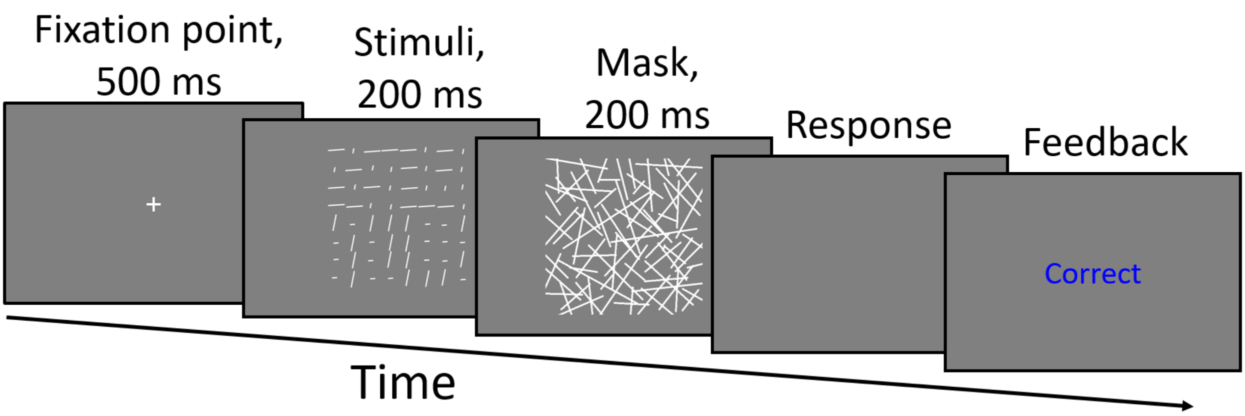


Fig. 2. The time course of a trial in Experiment 1.

Design and data analysis. In this experiment, we used a 2 (Segmentability: smooth vs. sharp) × 9 (Line-length correlation (as measured on the top half of a display): -1.00, -0.75, -0.50, -0.25, 0.00, 0.25, 0.50, 0.75, 1.00) within-subject design. 100 trials were presented within each factorial combination, so the total number of trials was 1,800 per observer.

To analyze the discrimination performance, we built psychometric functions for each observer and each segmentability condition. The psychometric functions plotted the probability of an "up" response *p* (Up) as a function of the length-orientation correlation in the top half of the display. Since some observers were instructed to search for a negative correlation, we inverted their original probabilities of "up" responses (1 – *p* (Up)). This transformation was done in order to get a standard shape for the psychometric function, where the growth of the response probability is expected with the growth of the stimulus (here, with a positive correlation). Therefore, values of *p* (Up), *unitized* across observers, were eventually used for building the psychometric functions.

The shapes and parameters of the psychometric functions were fit using normal cumulative density functions (CDF). Two parameters characterizing performance were extracted from each fitted function: μ (the mean of the normal distribution) and σ^2 (the variance of the normal distribution). The latter parameter was of principal significance, as it characterizes the discriminability of the stimulus: The smaller σ^2 we observe in a psychometric function, the better the observer discriminates the stimuli.

Results and discussion

For each participant, σ^2 for the smooth condition was greater than σ^2 for sharp one (observer A.Y.: 0.955 vs. 0.436; observer I.U.: 0.814 vs. 0.364; observer L.D.: 1.227 vs. 0.534; observer V.K.: 2.478 vs. 0.548; observer Y.S.: 1.099 vs 0.538). The graphical interpretation of σ^2 is the slope of the psychometric functions (CDF), with steeper slopes corresponding to a smaller σ^2 (Figure 3). We can conclude, therefore, that all five observers showed better discrimination of length-orientation correlations in "sharp" compared to "smooth" displays.

Two results of this experiment are important. First, we showed that observers were able to discriminate between different degrees of feature correlations, even in a short presentation with backward masking. The discriminability of correlated-feature patterns steadily grows with the physical contrast between their correlational structures (cf. Rensink, 2017; Rensink & Baldridge, 2010). Second and most important, segmentability makes correlation discrimination easier. The relatively steep slopes of the psychometric functions for the "sharp" condition (Figure 3) suggest a greater increment in sensitivity per unit of the physical feature correlation. In other words, any change in the correlational structure of the pattern is noticed with higher probability in "sharp" compared to "smooth" displays.

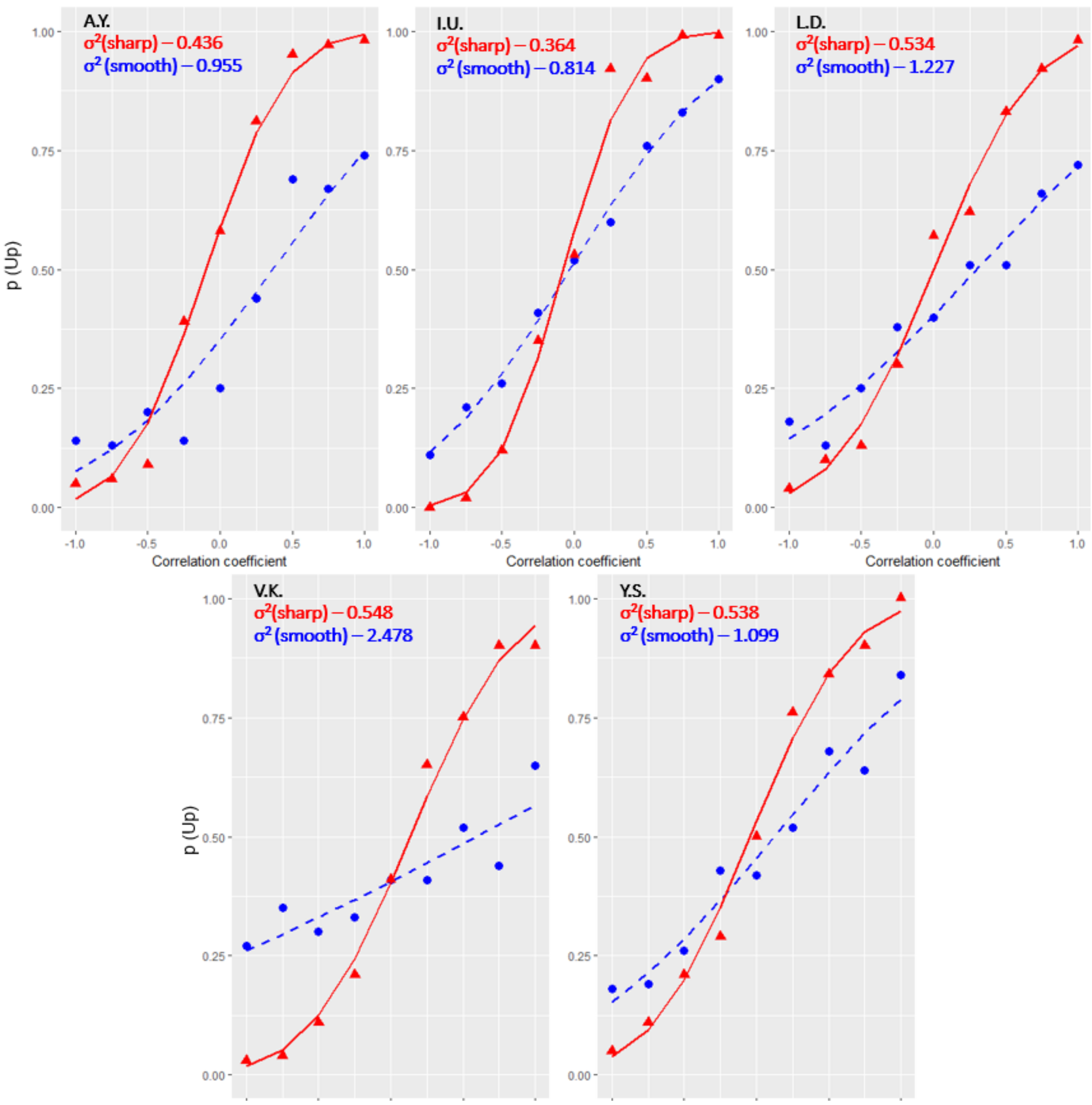


Fig.3. Psychometric functions for all observers plotting the unitized probability of responding "Up" as a function of length-orientation correlation in the top half of the display.

To summarize, Experiment 1 showed that the perceived feature correlation in conjunction-defined displays is a monotonic function of physical correlation and is improved by segmentability. In the next experiment, we test how this correlational structure is sampled over time and how conjunction-based segmentation arises from the segmentability of each feature dimension.

Experiment 2

In Experiment 1, we used a brief stimulus presentation (200 ms) to reflect the early (preattentive) processing of texture (Rosenholtz, 2000, 2014) or ensemble (Chong & Treisman, 2003; Whiting & Oriet, 2011) statistics. This is typically characterized by a widespread and rapid sampling of the visual field (Treisman, 2006). Our expert observers could achieve high accuracy with perfectly correlated (length-orientation correlation equal to -1 or 1) "sharp" displays, but their performance with "smooth" displays was relatively poor (Figure 3). It seems, therefore, that within the window of early processing, sampling of feature correlation is not very efficient, but segmentability somehow increases its efficiency. How does this sampling take place then? In previous research on ensemble summary statistics, a limited-capacity mechanism of focused attention is discussed as a candidate (Allik et al., 2013; Gorea, Belkoura, & Solomon, 2014; Myczek & Simons, 2008), although this mechanism is debated (Alvarez, 2011; Ariely, 2008; Chong, Joo, Emmanouil, & Treisman, 2008; Utochkin & Tiurina, 2014). If this is the case, then correlation discrimination should benefit from longer stimulus presentation, but this benefit should be observed later than 200 ms, as the attentional sampling rate is supposed to be no more than ~1-5 samples a second (Gorea et al., 2014; see also Wolfe, Alvarez, & Horowitz, 2000, for limiting conditions). Other studies demonstrated that faster sampling can take place in the accumulation of ensemble statistics (Chong & Treisman, 2003; Whiting & Oriet, 2011); these data suggest that fast non-attentional (or preattentive) sampling can exist. In Experiment 2, we examined several stimulus presentation durations (from 100 to 500 ms) to test whether correlation discrimination benefits from fast sampling, slow sampling, or both. Combining this with segmentability manipulation, this allowed us to see how the early ensemble-based segmentation of conjunctions occurs.

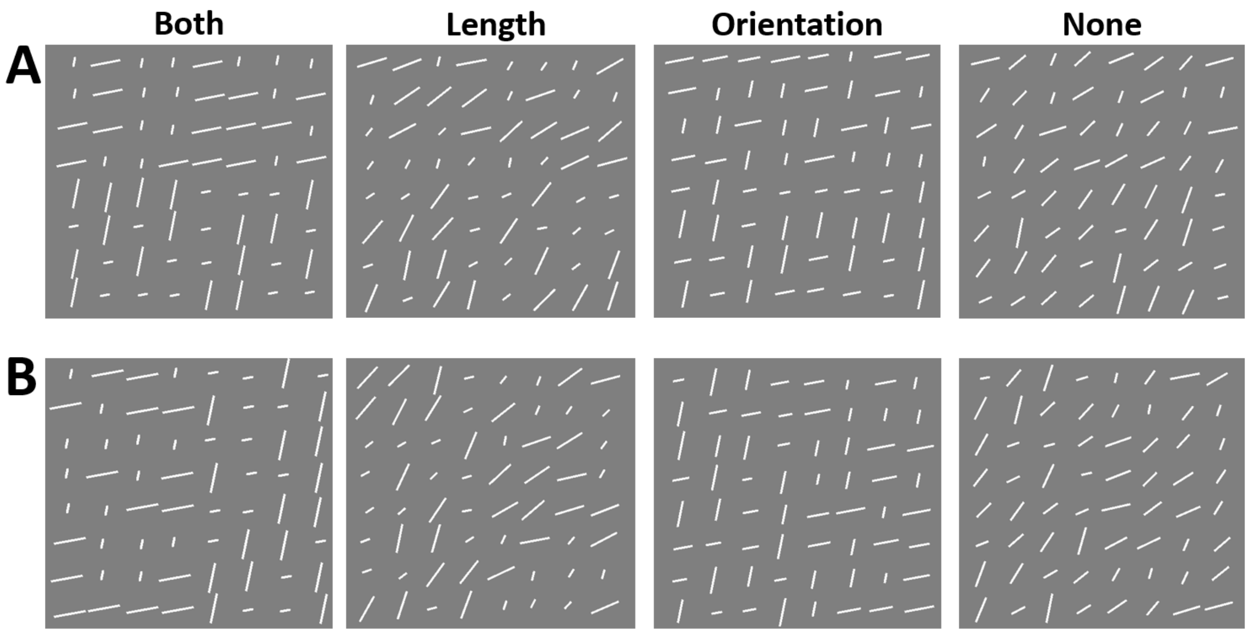
The second question we wanted to answer in Experiment 2 was how the conjunction segmentability, we observed in Experiment 1, originates from the more basic segmentability properties of separate feature distributions. While in Experiment 1 the distributions were either both smooth or both sharp, in Experiment 2 we manipulated the segmentability of the distributions orthogonally. Can observers better discriminate correlations if one dimension (e.g., length) is segmentable and the other one (e.g., orientation) is not? In other words, if they can easily select a well-segmentable subset (e.g., long lines), can they say "long-steeper" lines from "long-flatter" lines more easily, even when orientations are not segmentable? Or must both the length and the orientations be distributed sharply in order to provide better discrimination?

We used a slightly different psychophysical paradigm and a sample of observers in Experiment 2. As the correlation contrast between two textural patches was now fixed, we did not model psychometric functions. Instead, our paradigm was intended for obtaining the signal detection theory sensitivity rate d'. We also used a sample of naive observers instead of expert observers. This allowed us to collect a number of observations of sufficient statistical power (as d', unlike psychometric functions, is less informative in individual analysis) and generalize our findings to a broader population.

Methods

Participants. In total, 21 undergraduate students at the Higher School of Economics participated in Experiment 2 for extra course credits (all females, aged from 18 to 21 years, median age was 20 years). All reported having normal or correct-to-normal visual acuity and no experience of neurological problems. At the beginning of experiment, the participants signed an informed consent.

Apparatus and stimuli. We used the same apparatus as in the Experiment 1. Stimulation was also similar in terms of the spatial arrangement of line elements, their length and orientation ranges, and how segmentability was defined within each feature dimension. However, there were three important differences. First, in some displays the correlational structures could differ between the top and the bottom halves (as in Experiment 1), but in other displays it could differ between the left and the right halves (Figures 4A and 4B). This provided either a horizontal, or a vertical boundary between two correlation-defined sets (or textures). Second, we used a fixed correlational contrast between the halves, r = 1.00 vs. r = –1.00. The positive and negative correlations evenly alternated between the halves from trial to trial. Third, the segmentability of length and orientation distributions varied orthogonally from display to display. That is, the distributions could be both smooth, both sharp, length sharp with orientation smooth, and length smooth with orientation sharp. These four segmentability conditions were called, respectively, *both* segmentable, *none* segmentable, *length* segmentable, and *orientation* segmentable (Figure



4).

Fig. 4. Example stimuli used in Experiment 2 with four segmentability conditions. Panel (A) shows horizontal boundaries and panel (B) shows vertical boundaries between the textures.

Design and procedure. The participants were seated approximately 50 cm from the screen. Their task was to determine whether a boundary between the half-sets with opposite length-orientation correlations is horizontal or vertical. The concept of correlation was explained as “the longer the line, the steeper its slope” (and vice versa), and also as prevalence of “longer-steeper and shorter-flatter” (and vice versa). For visualization purposes, the experimenter drew an example set during the explanation, pointing to the differences in line attributes between the patches with positively and negatively correlated features. This provided a clear understanding of the task for the participants.

Each trial started with a presentation of a fixation point for 500 ms. The sample display was shown for 100, 200, 350, or 500 ms and followed by a mask for 200 ms. After that participants had to report whether the boundary between the patches with opposing correlations had been horizontal or vertical. They used “left” and “up” arrow keys, respectively. Responses were followed by a feedback informing the observers whether answer had been correct or not. In the practice session at the beginning of the experiment, feedback was accompanied by the repeated presentation of the sample stimulus for unlimited time, with horizontal or vertical color line markers indicating the beginning and the end of the boundary (if the response had been correct the markers were blue, otherwise they were red). A next trial started by pressing the SPACE bar, so participants could progress at a comfortable pace and take a rest whenever they wanted.

Design and data analysis. In this experiment, we used a 4 (Stimulus duration: 100, 200, 350, and 500 ms) × 4 (Segmentability: both, length, orientation, and none) within-subject design. 40 trials was presented within each factorial combination (20 with horizontal and 20 with vertical boundaries), so the total number of trials was 640 per observer.

The sensitivity index d' was calculated from the data, using a formula for two-alternative force choices (Stanislaw & Todorow, 1999). A 4 × 4 repeated-measure ANOVA (the model corresponding to the design) was then applied to d'. We implemented the standard significance tests and calculated Bayes factors for our ANOVA model. In the Bayesian statistical inference, the Bayes factor (BF_{10}) is the odds showing the relative likelihood of H_1 compared to H_0 given the data. The Bayes factors were calculated in JASP statistical software (JASP Team, 2017). The width r of a prior Cauchy distribution of effect sizes was set at 0.5, following the default settings recommended by Wagenmakers and colleagues (Wagenmakers et al. 2017) and JASP Team (2017) for fixed effects models. Jeffreys' scale (Jeffreys, 1961) with Kass and Raftery's adjustment (Kass & Raftery, 1995) was used to interpret the Bayes factors. Here $1 < BF_{10} < 3$ is evidence for H_1 that is "not worth more than a bare mention" (Kass & Raftery, 1995, p. 777), $3 < BF_{10} < 20$ is positive evidence for H_1, $20 < BF_{10} < 150$ is strong evidence for H_1, and $BF_{10} > 150$ is very strong evidence for H_1. For post hoc pairwise comparisons, Bayes factors were calculated using the Bayesian t-test (Rouder et al., 2009) with a prior width set at r = 0.707, which is recommended as the default value for this test (JASP Team, 2017; Wagenmakers et al., 2017).

Results and discussion

We found a strong effect of stimulus duration on d' ($F(3,60) = 16.871$, $p < .001$, $\eta^2_p = 0.458$, $BF_{10} = 4.86\times10^5$). Post hoc t-tests showed that d' in 100 ms was lower than in the rest of the durations (t's > 4.44, p's < 0.001, Bonferroni corrected $\alpha = 0.008$, Cohen's d's > 0.48, BF_{10}'s > 649), while other durations showed no difference from each other (t's < 1.94, p's > 0.05, Bonferroni corrected $\alpha = 0.008$, d's < 0.23, BF_{10}'s < 0.72). The effect of segmentability was also very strong ($F(3,60) = 26.829$, $p < 0.001$, $\eta^2_p = 0.573$, $BF_{10} = 1.04\times10^{16}$). This was shown by the substantially greater d' in the "both" condition compared to the other three conditions (t's > 7.52, p's < 0.001, Bonferroni corrected $\alpha = 0.008$, d's > 0.82, BF_{10}'s > 1.6×10^8), while the rest of the conditions showed no difference from each other (t's < 1.33, p's > 0.19, Bonferroni corrected $\alpha = 0.008$, d's < 0.15, BF_{10}'s < 0.29). We did not find any evidence of an effect of stimulus duration × segmentability on d' ($F(9,180) = 1.499$, $p = 0.151$, $\eta^2_p = 0.070$, $BF_{12} = 0.81$, where BF_{12} is a ratio of BF_{10} for a model taking only the main effects into account to BF_{10} for a model taking the main effects and interaction into account). Additionally, we compared the observed d' for all factorial combinations with $d' = 0$ to estimate whether performance was better than chance in each condition. All of d' were found to be greater than 0 (t's > 77.8, p's < 0.001, Bonferroni corrected $\alpha = 0.003$, d's > 3.9, BF_{10}'s > 6.87×10^{10}), suggesting observers always were better than chance. The results are shown in Figure 5.

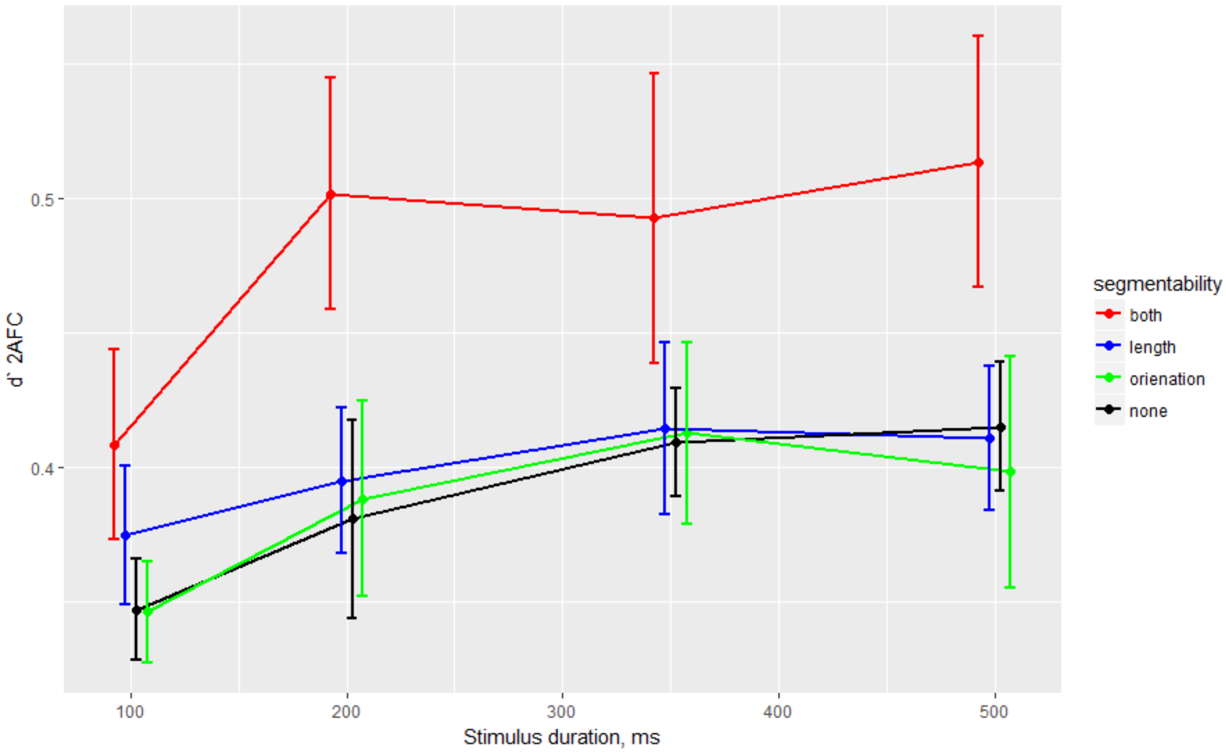


Fig. 5. Texture discriminability (d') as a function of presentation duration and segmentability (experiment 2).

Our results show that observers were better at discriminating two sets (textures) with different correlations only when both correlated dimensions had sharp feature distributions. This finding replicates the advantage of segmentability observed in Experiment 1, though in a slightly different paradigm. Importantly, the two newly introduced conditions, where only one dimension had a sharp distribution, failed to improve correlation discrimination as compared to the "none" condition. This suggests that conjunction segmentability arises from all contributing features being segmentable. In the next section, a more detailed explanation will be provided, based on the results of Experiments 1 and 2, for why segmentability rather than other statistical properties of the distributions is crucial for correlation perception.

The second important effect shown in Experiment 2 is the effect of stimulus duration. We showed that the most dramatic increment in performance tended to occur within the early temporal window (from 100 to 200 ms) rather than within the late temporal window (from 350 to 500 ms). This finding is consistent with the idea of relatively fast sampling of correlational statistics, whose speed is comparable to the sampling speed of averaging (Chong & Treisman, 2003; Whiting & Oriet, 2011). In contrast, our data show no evidence for any benefit from slow sampling associated with focused attention (Allik et al., 2013; Gorea et al., 2014; Myczek & Simons, 2008). Moreover, the additive main effects of segmentability and stimulus duration suggest that conjunction segmentation can start at least as early as 100 ms.

General discussion

Previous research demonstrated that the distribution of individual features along a sensory dimension can provide an efficient basis for segmenting multiple objects based on that dimension (Treue, Hol, & Rauber, 2001; Utochkin & Yurevich, 2016). Substantial discontinuities between the clusters of similar features provide the distribution property that we called “segmentability” (Utochkin & Yurevich, 2016) and which can aid primitive but rapid categorization (Utochkin, 2015). In this study, we demonstrated that segmentability can work beyond separate dimensions. We found evidence that rapid segmentation of multiple conjunctions can emerge from good segmentability of the underlying feature dimensions, and that this segmentation supports discrimination between sets with similarly distributed but differently correlated (or bound) features.

Segmentation vs. other ensemble-based algorithms

Our manipulations with the shapes of length and orientation distributions made some other statistical properties of these distributions change as well. Could these changes, rather than segmentability per se, account for our results? Below, we consider these alternative accounts and show why our data are best explained by segmentability.

Variance is the first obvious ensemble parameter that changes with the shape of distribution. It is easy to see that the distributions with only extreme feature values ("sharp" condition in our experiments) have greater total variance than the distributions with extreme and medium feature values ("smooth" condition in our experiments). However, it seems very unlikely that the total variance could explain the facilitatory effects of sharp distributions. Previous research has convincingly shown that higher feature variance reduces the perceived contrast between the summary statistics of different sets (Callaghan, 1984, 1989; Corbett et al., 2012; Fouriezos et al., 2008; Im & Halberda, 2013; Maule & Franklin, 2015; Rosenholtz, 2000; Utochkin & Tiurina, 2014). In our experiments, the sharp and, thus, more variable distributions enhanced rather than reduced discriminability. We, therefore, ruled out total variance as an explanation for our data pattern. The shape itself seems to be a preferred explanation at this point.

Another possible strategy is selecting a subset of items whose features are located roughly close to one side of the distribution of the first dimension ("half-splitting"). Within this subset, a few samples can be tested for the average feature along a second dimension. For example, an observer can attend to a subset of longer lines and establish which samples of this subset are steeper or flatter on average. This might give the impression of differences between patches without computing the full correlations.

The idea of half-splitting can explain the observed enhancement effect of segmentability. As sharp distributions consist of clearly shaped peaks, determining which feature represents which side of the distribution is not hard. By contrast, smooth displays imply a wide single-peak distribution where half-sets are harder to determine. One reason is that, without having a gap between the extremes, features from the opposite sides of a smooth distribution can be confused. This is especially true for the features lying closer to the center of the distribution. To avoid this, an observer could ignore the medium features and focus on the tail of the distribution (e.g. on the longest lines only). But due to the smoothness of inter-feature transition (that is, the lack of segmentability), these "tail-most" elements are not salient enough to be quickly and accurately distinguished from the medium. As a result, the half-sample of the smooth distribution is imperfect and the corresponding correlation judgments are less accurate.

The half-split strategy, described above, assumes that only one dimension should be segmentable. If a highly segmentable subset of long lines is selected, then the orientations can be compared based on average, so its segmentability should not play a substantial role. In Experiment 2, we found that segmentability enhances texture discrimination only when both dimensions were represented by the sharp distributions. The critical question is whether this is explained by segmentation or better average comparison within the second dimension when it has a sharp distribution. To illustrate this point, consider an example stimulus with a correlation contrast "$r = 1.00$ vs. $r = -1.00$" from our experiments. If the length is segmentable and a half-set of long lines is selected, only orientation contrasts should be found between the regions for establishing correlation differences. If orientations are also distributed sharply, their segmentability would provide good discrimination between "long-steep" and "long-flat" lines; if they are distributed smoothly, it is more difficult to distinguish between steep and flat lines, and this predicts poorer performance. However, the discriminability between the samples of orientations within the subset of long lines can be explained by the differences in the means and variances. In sharp distributions consisting of 11º and 86º elements, the physical orientational mean difference between the length-limited sub-textures was 75º and the standard deviation of each sub-texture was 0º. In the smooth condition, the mean difference between the sub-textures was 40º, and the standard deviation of each sub-texture was about 12º. For a symmetrical case (orientation is segmented), the mean length differences between the orientation-limited textures was 2.1º (3.3 times) in the sharp and 1.1º (1.8 times) in the smooth conditions; the standard deviations of the sub-textures were 0º and 0.35º, respectively. From the previous work, we know that a smaller mean difference combined with a greater standard deviation within the textures leads to a decrement in average discrimination between the sets (Corbett et al., 2012; Fouriezos et al., 2008; Im & Halberda, 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014).

Having said that, the data from the previous work show that, even with smaller mean differences and greater standard deviations, the contrasts between sub-textures in our experiments were sufficient to provide a mean discrimination above the threshold. In the tasks requiring discriminations between mean orientations (Rosenholtz, 2000), mean lengths (Fouriezos et al., 2008), or mean dot diameters (Chong & Treisman, 2003, 2005b; Im & Halberda, 2013) between two sets, observers were more accurate with lower mean difference-to-variance ratios than our participants, when working with one sharp and one smooth distribution. If our observers compared the averages along a second dimension within a half-sample along a first dimension, we would have expected better performance in these conditions compared to performance in the “none” condition (when both distributions are non-segmentable and half-splitting is not reliable at all). One could argue that the half-set distribution can be permeable to the features from another half, which increases actual variance and decreases the actual mean difference. But this explanation seems hardly distinguishable from our previously adopted explanation in terms of segmentability. If half-set statistics of a smooth distribution are susceptible to the influence of another half-set more than the sharp half-set (where the other half-set consists of opposing extreme and thus potentially even more biasing features), it indicates a problem with establishing a boundary between the half-sets. We conclude, therefore, that our finding that only both sharp distributions enhance the discrimination between the correlational structures of two textures, is explained by double segmentation and the formation of categorical classes of conjunctions.

The time course of sampling

Our data show that correlation statistics (statistics of feature conjunctions) are less discriminable than single feature statistics, as previously reported (e.g. Rosenholtz, 2000). However, Experiment 2 shows that people do somehow estimate feature correlation above chance rapidly, even in the hardest condition with both length and orientations distributed smoothly. Note that at the shortest duration of 100 ms the accuracy is already non-zero, suggesting that some information about feature correlation is available early. At the same time, the increase in the presentation time beyond 200 ms did not have an effect on the accuracy of correlation estimation. This fact leads to two conclusions. Firstly, observers did not use sampling strategies based on focused attention (cf. Allik et al., 2013; Gorea et al., 2014; Myczek & Simons, 2008). If this was the case, participants would have acquired more precise information via collecting more local samples with time, so that *d*` should have grown. Second, as our task requires the binding of these features (or referencing between the features) and *d*` is always above 0, we can conclude that some coarse binding can take place without attention (e.g., Rosenholtz et al., 2012a). We do not claim, however, that our results dismiss the role of attention completely in object binding and recognition (Cave & Wolfe, 1999; Wolfe, Võ, Evans, & Greene, 2011): Our relatively low *d′* values show that the binding processes were very imperfect when lacking focused attention. It is possible that this coarse imperfect binding is sufficient to extract some statistics over large textured regions but insufficient for more precise recognition and localization of particular feature conjunctions (Balas et al., 2009; Wolfe, 1992).

Despite having no evidence for slow attentional sampling, we observed increasing sampling efficiency within the “preattentive” window, between 100 and 200 ms. Although the correlation between two dimensions is a more complex statistic than the average along a single dimension, the time course of accumulating the information about correlations is comparable with the time course of visual averaging (Whiting & Oriet, 2011). Importantly, the effect of feature segmentability was observed very early: there was a significant benefit from both sharply distributed features at 100 ms. In our opinion, this early segmentation mediates sampling and explains the gain from the sharp distributions. When the items are well segmented into classes, observers become more effective in drawing proper samples. In contrast, when distributions are smooth, observers are worse at using this strategy, because any pair of samples for comparison is more similar.

Implications for the theory of vision

Our results are important for understanding how discrete groups of objects can be coarsely identified for a more rigorous analysis. Although the features of real-world objects have some natural variation along continuous dimensions (such as size, orientation, color wavelength, etc), these continuously distributed features can be discretized based on the shape of distribution. In our experiments (Experiment 2 in particular), we demonstrated how the discrete classes of items arise from statistical properties of more elementary distributions, each corresponding to one variable sensory dimension. Segmentability, an earlier discovered emergent property of the single feature distribution (Utochkin & Yurevich, 2016), seems to provide a basis for discretization and coarse categorization (Utochkin, 2015). Our new data show that segmentability is a requirement for all variable features in order to support discrimination between the classes of feature conjunctions. In our experiments, the categorical differences supported by both segmentable lengths and orientations provide the facilitation of texture discrimination via access to categorically distinct subsets.

The global and early segmentation of a scene can play an important role in the further deployment of attention and scene recognition (Wolfe et al., 2011). Our results provide new information about the boundary conditions for such segmentation and attentional control. According to the guided search model (Wolfe, 1994; Wolfe et al.,1989), observers can limit the number of inspected items to the preattentively segmented part of the set sharing common features with the target template. For example, when searching a red vertical line among multiple red horizontal and green vertical lines, the observer can give more weight to all red items and search for the only vertical item among them, which is the target (Friedman-Hill & Wolfe, 1995;

Wolfe et al., 1989). A similar mechanism was reported for binocular depth as a segmentation cue in search of depth-color or depth-motion conjunctions (Nakayama & Martini, 2011; Nakayama & Silverman, 1986). This strategy leads to a much more efficient conjunction search than that predicted by the purely serial deployment of attention (Treisman & Gelade, 1980). Our results show that the ability to select subsets for a guided conjunction search should depend on the segmentability of conjunction-defining dimensions. For example, when an observer attends to the items of a certain color, the color of a relevant subset should be segmentable from the irrelevant one. A less trivial point following from our results is that the second dimension should be segmentable, too. That is, even when color is segmentable enough for the selection of a relevant subset, the guided visual search without orientation being also segmentable will be less efficient. This prediction was not tested in the present study but can be tested in further experiments.

Conclusion

The results of our study advance our understanding of how the visual system splits poorly organized sets of multiple items (variable in attributes and intermixed in the space, which often corresponds to the natural organization of the real world) into groups representing categorically different classes of objects. Ensemble summary statistics, proven as an efficient tool to organize visual cognition in many different ways, seem to guide the massive transformation of continuously distributed visual features into discrete categories of objects. Here we showed that segmenability, a previously described emergent property associated with the shape of feature distribution, plays an important role in the categorization of objects whose categorical differences are defined by particular conjunctions of more simple features.

References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83,* 25-39. doi:10.1016/j.visres.2013.02.018

Alvarez, G. A. & Oliva, A. (2008). The representation of simple ensemble features outside the focus of attention. *Psychological Science, 19(4),* 392-398. doi:10.1111/j.1467-9280.2008.02098.x.

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

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162. doi:10.1111/1467-9280.00327.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception and Psychophysics, 70 (7),* 1325–1326. doi:10.3758/PP.70.7.1325.

Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision, 8(4)*:9, 1–22. doi: 10.1167/8.4.9.

Balas, B., Nakano, L., & Rosenholtz, R. A. (2009). Summary-statistical model of peripheral vision explains visual crowding. *Journal of Vision, 9(12)*:13, 1–9. doi:10.1167/9.12.13.

Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, *59* (2), 171-185.

Bauer, B. (2017). Perceptual averaging of line length: Effects of concurrent digit memory load. *Attention, Perception, and Psychophysics, 2017 August 4.* [Epub ahead of print]. doi:10.3758/s13414-017-1388-4.

Burr, D. & Ross, J. (2008). A visual sense of number. *Current Biology, 18(6)*, 425–428. doi:10.1016/j.cub.2008.02.052.

Callaghan, T. C. (1984). Dimensional interaction of hue and brightness in preattentive field segregation. *Perception and Psychophysics, 36(1),* 25-34.

Callaghan, T. C. (1989). Interference and dominance in texture segregation: hue, geometric form, and line orientation. *Perception and Psychophysics*, *46(4),* 299-311.

Cave K.R. & Wolfe, J.M. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron,* 2*4(1)*, 11-17.

Chong, S. C. & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404. doi:10.1016/S0042-6989(02)00596-5.

Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception and Psychophysics, 70,* 1327-1334. doi:10.3758/PP.70.7.1327.

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

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

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences, 20*, 324-335. doi: 10.1016/j.tics.2016.03.006.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185. doi: 10.1017/S0140525X01003922.

Dakin, S. C. & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37,* 3181-3192.

De Gardelle, V. & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, (32), 13341-13346.

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences, 16*(6), 317–321. doi:10.1016/j.tics.2012.04.007.

Duncan, J. & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Emmanouil, T. A. & Treisman A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, 70 (6), 946–954.

Epstein, M. L. & Emmanouil, T. A. (2017). Ensemble coding remains accurate under object and spatial visual working memory load. *Attention, Perception and Psychophysics, 2017* Jun 9. [Epub ahead of print]. doi:10.3758/s13414-017-1353-2.

Found, A. (1998). Parallel coding of conjunctions in visual search. *Perception & Psychophysics, 60(7),* 1117-1127.

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception and Psychophysics, 70, 3,* 456–464. doi:10.3758/PP.70.3.456.

Friedman-Hill, S. R, & Wolfe, J.M. (1995) Second-Order Parallel Processing: Visual Search for the Odd Item in a Subset. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3): 531-551.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision, 14(9):*22, 1-14. doi: 10.1167/14.9.22.

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

Haberman, J. & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression.*Attention, Perception, and Psychophysics. 72(7),* 1825-1838. doi:10.3758/APP.72.7.1825.

Haberman, J. & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe and L. Robertson (Eds.), *From Perception to Consciousness: Searching with Anne Treisman.* Oxford University Press, 339-349.

Huang, L. (2015). Statistical properties demand as much attention as object features. *PLoS ONE, 10(8),* e0131191. doi: 10.1371/journal. pone.0131191.

Humphreys, G. W. & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25,* 43-110. doi: 10.1006/cogp.1993.1002.

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, and Psychophysics, 75(2),* 278–286. doi:10.3758/s13414-012-0399-4.

Jackson-Nielsen, M., Cohen, M.A., & Pitts, M. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition,* 48, 149-160. doi:10.1016/j.concog.2016.11.007.

JASP Team (2017). JASP (Version 0.8.2) [Computer software]

Jeffreys, H. (1961). Theory of Probability, 3rd ed. Oxford Classic Texts in the *Physical Sciences*. Oxford Univ. Press, Oxford. MR1647885.

Kass, R.E., Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association. 90(430)*, 773-795.

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281. doi: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. doi:10.1167/15.4.6.

Müller, H. J., Humphreys, G. W., & Donnelly, N. (1994). Search via recursive rejection (SERR): Visual search for single and dual form-conjunction targets. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 235-258. doi: 10.1037/0096-1523.20.2.235.

Myczek, K. & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception and Psychophysics, 70,* 772–788.

Nakayama, K. & Martini, P. (2011). Situating visual search. *Vision Research, 51 (13),* 1526-1537. Doi: 10.1016/j.visres.2010.09.003.

Nakayama, K. & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320(6059),* 264. doi:10.1038/320264a0.

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, 1–14. doi:10.1167/15.4.3.

Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion, and color cues. *Vision Research, 33,* 1937-1958.

Nothdurft, H. T. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception and Psychophysics, 52,* 355-375.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227-1268.

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

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162(1-2),* 8-13, doi:10.1016/j.jneumeth.2006.11.017.

Pylyshyn, Z. & Storm, R. (1988).Tracking multiple independent objects: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179–197.

Rensink, R. A. & Baldridge, G. (2010). The perception of correlation in scatterplots. *Computer Graphics Forum*, *29,* 1203-1210.

Rensink, R. A. (2017). The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, **24**: 776-797. doi: 10.3758/s13423-016-1174-7.

Robitaille, N. & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11(12),* 18. doi:10.1167/11.12.18.

Rosenholtz, R. (2000). Significantly different textures: A computational model of pre-attentive texture segmentation. In Vernon D. (Ed.) *Proceedings of the European Conference on Computer Vision (ECCV '00), LNCS 1843,* 197–211.

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 and Performance, 27(4),* 985-999. doi:10.1037//0096-1523.27.4.985.

Rosenholtz, R. (2014). Texture perception. In Wagemans J. (Ed.) *Oxford handbook of perceptual organization* (pp 167–186). Oxford, UK: Oxford University Press.

Rosenholtz, R., Huang, J., & Ehinger, K. (2012a). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, *3(13)*, doi:10.3389/fpsyg.2012.00013.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012b). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12 (4)*: 39 1–17, doi:10.1167/12.4.14.

Rouder, J. N., Speckman P. L., Sun D., Morey R. D., & Iverson G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16(2)*, 225-237. doi:10.3758/PBR.16.2.225.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition*, *80*, 1–46. doi: 10.1016/S0010-0277(00)00152-9.

Simons, D. J. & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception and Psychophysics, 70*, 1335–1336. doi:10.3758/PP.70.7.1335.

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31*, 137-149. doi:10.3758/BF03207704.

Treisman, A. (1999). Solution to the binding problem: Progress through controversy and convergence, *Neuron 24*, 105–110.

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

Treisman, A. M. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion – physiology and psychophysics. *Nature Neuroscience, 3,* 270–276. doi:10.1038/72985.

Tsotsos, J. K. (1987). A Complexity Level analysis of immediate vision. *International Journal of Computer Vision* ***1****(4),* 303–320.

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. doi: 10.1016/j.actpsy.2013.11.012.

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. doi:10.1037/xhp0000203.

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization . *Journal of Vision, 15(4):8,* 1-14. doi:10.1167/15.4.8.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., et al. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 2017 July 6*. [Epub ahead of print]. doi: 10.3758/s13423-017-1323-7.

Watamaniuk, S. N. J. & Duchon, A. (1992). The Human Visual-System Averages Speed Information. *Vision Research, 32(5),* 931-941.

Whiting, B. F. & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review, 18(3)*, 484–489.

Wolfe, J. M. (1992). “Effortless” texture segmentation and “parallel” visual search are not the same thing. *Vision Research 32*, 757–763.

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

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 and Performance, 15*(3), 419-433.

Wolfe, J. M., Vo, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cogniive Sciences*, 15(2), 77-84.

Wolfe, J.M., Alvaraz, G.A., & Horowitz, T.S. (2000). Attention is Fast but Volition is Slow. *Nature, 406,* 691.

Yamanashi-Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications, 7,* 13186. doi: 10.1038/ncomms13186.

Yantis, S. (2014). *Sensation and Perception*. New York, NY: Worth Publishers.

Igor S. Utochkin
National Research University Higher School of Economics (Moscow, Russia). “Laboratory for Cognitive Research”. Principal investigator;
E-mail: iutochkin@hse.ru, Tel. +7 (495) 772-95-90*23-114