Acta Psychologica 213 (2021) 103238

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Size-distance rescaling in the ensemble representation of range: Study with binocular and monocular cues

Yuri A. Markov *, Natalia A. Tiurina

National Research University Higher School of Economics, Russia

ARTICLE INFO

Keywords:
Ensemble summary statistics
Feature binding
Size constancy

ABSTRACT

According to numerous studies observers can rapidly and precisely evaluate mean or range of the set. Recent studies have shown that the mean size estimated based on sizes of objects rescaled to their distances (Tiurina & Utochkin, 2019). In the current study, we directly tested this rescaling mechanism on the perception of range using binocular and monocular cues.

In Experiment 1, a sample set of circles with different angular sizes and in different apparent distances were stereoscopically presented. Participants had to adjust the range of the test set to match the range of the sample set. The main manipulation was the size-distance correlation for sample and test sets: in negative size-distance correlation, the apparent range had to decrease, while in positive correlation - increase. We found the highest underestimation in the condition with the negative sample correlation and positive test correlation, which could be explained only if ensemble summary statistics were estimated after the item's rescaling.

In Experiment 2, we used Ponzo-like illusion and spatial positions as a depth cue. Sets were presented with positive, negative or without size-distance correlation on a grey background or the background with Ponzo-like illusion. We found that the range was underestimated in negative correlation and overestimated in positive correlation.

Thus, items of ensemble could be automatically rescaled according to their distance, based on both binocular and monocular cues, and ensemble summary statistics estimation is based on perceived sizes.

1. Introduction

The perceived world is complex and consists of a vast amount of textures and colors, which form different objects and create visual scenes. Our perception system is highly limited, it can only process a small number of objects (Cowan, 2001; Luck & Vogel, 1997; Miller, 1956) and the capacity of our attention is limited too (Mack & Rock, 1998; Pylyshyn & Storm, 1988; Simons & Chabris, 1999). However, we perceive the world as richly detailed and stable (Cohen, Dennett, & Kanwisher, 2016).

Some explanations of how our visual system overcomes these limitations (Alvarez, 2011) are theories suggesting that our visual system estimates summary statistics (or ensemble summary statistics). Ariely first showed (Ariely, 2001) that observers had a better representation of the mean size of the ensemble, rather than individual representations of each object. The visual system could rapidly and automatically estimate statistics for the set of objects by one feature without having precise representation of each object (Chong & Treisman, 2005; Whiting & Oriet, 2011). The ability to rapidly estimate different statistics were shown for the mean size (Ariely, 2001; Chong & Treisman, 2003), brightness (Bauer, 2009), orientation (Alvarez & Oliva, 2008; Attarha & Moore, 2015), speed (Watamaniuk & Duchon, 1992), even emotion or gender of a crowd of faces (Haberman & Whitney, 2007). Our visual system could not only estimate mean, but also range or variance (Dakin & Watt, 1997; Haberman, Lee, & Whitney, 2015; Jeong & Chong, 2020; Khvostov & Utochkin, 2019; Morgan, Chubb, & Solomon, 2008; Norman, Heywood, & Kentridge, 2015; Solomon, Morgan, & Chubb, 2011; Suárez-Pinilla, Seth, & Roseboom, 2018), and numerosity (Burr & Ross, 2008; Halberda, Sires, & Feigenson, 2006; Khvostov & Utochkin, 2019; Chong & Evans, 2011; Utochkin & Vostrikov, 2017).

However, most of the research is devoted to the study of the ensemble statistical representation based on a single visual dimension, which is not quite ecologically valid and does not fully correlate with the real conditions of perception: in real life we encounter different

* Corresponding author at: 101000, 4 Armyanskiy pereulok Building 2, Psychology Department, National Research University Higher School of Economics, Moscow, Russian Federation.

E-mail address: yuamarkov@gmail.com (Y.A. Markov).

https://doi.org/10.1016/j.actpsy.2020.103238
Received 22 March 2020; Received in revised form 8 October 2020; Accepted 9 December 2020
Available online 30 December 2020
0001-6918/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

dimensions. There are small number of studies that attempt to elucidate the mechanism of ensemble representation based on several visual dimensions (Emmanouil & Treisman, 2008; Huang, 2015; Im, Tiurina, & Utochkin, 2020; Leib et al., 2014; Tiurina & Utochkin, 2019; Utochkin, Khvostov, & Stakina, 2018). One part of these studies (Emmanouil & Treisman, 2008; Huang, 2015) investigated the role of attention divided between different dimensions in ensemble perception. Another part of these studies (Im et al., 2020; Utochkin et al., 2018) explored the segmentation and categorization of the set consisting of several visual feature distributions. Though, it is little known (Leib et al., 2014; Tiurina & Utochkin, 2019) about ensemble summary statistics evaluation with integration of several dimensions.

There is one dimension, which is commonly used in ensemble summary statistics studies (Whitney & Yamanashi Leib, 2018), and which, in the real-world, is represented by two separate features: retinal size and depth (or distance to the observer) - perceived size. From the perspective of perceptual constancy, our visual system can rescale retinal images to get information about physical real-world objects. According to the size constancy theories (Boring, 1940; Epstein, Park, & Casey, 1961; Holway & Boring, 1941; Kaufman et al., 2006; Sperandio & Chouinard, 2015) angular size could be correctly and efficiently rescaled based on various depth cues. Size distance invariance hypothesis proposes that "*visual angle of given size determines a unique ratio of apparent size to apparent distance*" (Boring, 1940; Epstein, Park, & Casey, 1961; Kilpatrick & Ittelson, 1953). Depth cues could be divided into three categories: oculomotor, binocular and monocular depth cues (Palmer, 1999). There are various types of binocular cues (binocular disparity (Deangelis, 2000), vertical disparity, shadow stereopsis, etc. (Puerta, 1989)) and monocular cues (perspective, relative size, familiar size, etc. (Palmer, 1999)). There are different mechanisms of size-distance rescaling based on binocular and monocular cues. They are processed via two distinct pathways (Goodale & Milner, 1992; Mishkin & Ungerleider, 1982): binocular cues (disparity) by dorsal stream (Backus, Fleet, Parker, & Heeger, 2001; Mon-Williams, Tresilian, McIntosh, & Milner, 2001; Qian & Yazdanbakhsh, 2015) and monocular cues (perspective, etc.) by ventral stream (Fang, Boyaci, Kersten, & Murray, 2008; Marotta, Behrmann, & Goodale, 1997; Mikellidou et al., 2016; Weidner et al., 2014; but see Welchman, Deubelius, Conrad, Bülthoff, and Kourtzi (2005) suggesting that perspective cue processing could involve a dorsal network). V1 plays an important role in depth perception in both binocular and monocular cues (Fang et al., 2008; Murray, Boyaci, & Kersten, 2006; Pooresmaeili, Arrighi, Biagi, & Morrone, 2013; Sperandio & Chouinard, 2015; Sperandio, Chouinard, & Goodale, 2012; Trotter, Celebrini, & Durand, 2004), activity of V1 neurons modulated by the perceived rather than the retinotopic sizes of objects. Mediation of size-distance scaling occurs in the primary visual cortex (V1): size tuned cells in V1 are modulated via top-down feedback from higher brain areas (Qian & Yazdanbakhsh, 2015; Sperandio & Chouinard, 2015). Feedback into V1 for binocular cues possibly comes from V3a, middle temporal visual area (MT) and lateral intraparietal cortex (LIP) (Gnadt & Mays, 1995; Qian & Yazdanbakhsh, 2015), while for monocular cues mainly from lateral occipital cortex (LOC) and parahippocampal place area (PPA) (Fang et al., 2008). Such a neural circuit provides a compensatory mechanism for maintaining size constancy by scaling retinal size to viewing distance (Qian & Petrov, 2016; Qian & Yazdanbakhsh, 2015).

These studies are usually investigated rescaling of one item (for review Sperandio & Chouinard, 2015), while feature integration theory (Treisman, 1996; Treisman & Gelade, 1980) suggests binding of several dimensions for multiple objects could be quite a complex process – the visual system needs to correctly perceive and connect each angular size with each distance (Tsotsos, 1988) and this process could require focused attention (Emmanouil & Treisman, 2008). However, there are studies suggesting that it is possible to bind features even without attention (Di Lollo, 2012; Rosenholtz, Huang, & Ehinger, 2012). Angular size and depth could influence the perceptual interpretation of each other – this is one of the examples of "conditional binding" (Treisman, 1996), thus this process could differ from other types of binding of other features. Nevertheless, it could still be a complex process, which could require some resources.

The recent study by Tiurina and Utochkin (2019) investigated how mean size is estimated, based on two dimensions (angular size and depth): is it based on "raw" representations (unbound) or on perceived (bound)? In Experiment 1 participants overestimated the mean of the set presented using a mirror stereoscope farther from the observer and underestimated the mean of the set presented closer to the observer. However, all items had the same depth and it is not obvious how observers bound features: every individual angular size with every individual depth or the mean size was rescaled to the ensemble depth. In Experiment 2, observers reported the mean size of the ensemble, where each item had individual depth and angular size. Key manipulation of Experiment 2 - the correlation between angular size and depth. In the positive correlation condition, small circles were presented closer to observers and larger circles were presented farther, and vice versa for the negative condition. If rescaling happens after mean estimation, when there will be no difference between these conditions, while if the mean size is estimated on perceived sizes, when the perceived range will be higher for the positive correlation (small circles will appear smaller and big circles appear bigger), than for the negative correlation. The main result of Experiment 2 demonstrated the smaller error of mean size estimation for the negative correlation, which is indirect demonstration that perceived range became narrower. So, according to results rescaling happens before averaging, and our visual system firstly bound information about angular size and depth, and only then estimated mean size.

Our current study elaborates Tiurina and Utochkin (2019) study. Here we directly tested the rescaling mechanism on the perception of range, using both binocular and monocular cues. As we discussed previously processing of binocular and monocular cues are highly different (McKee & Taylor, 2010), thus, it is necessary to demonstrate that multiple items also could be rescaled by monocular cues, since its processing happens in the LOC and PPA, which are also associated with processing of ensembles (Cant & Xu, 2012, 2015, 2017). We investigated how range is represented by several features. Though range is connected with average estimation (Fouriezos, Rubenfeld, & Capstick, 2008; Im & Halberda, 2013; Marchant, Simons, & de Fockert, 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014), still range representation is independent from mean representation and could be calculated in parallel with mean (Khvostov & Utochkin, 2019; Yang, Tokita, & Ishiguchi, 2018; but see Jeong & Chong, 2020). We also expanded and replicated previous findings (Tiurina & Utochkin, 2019). In two experiments we asked our participants to report the range of the presented ensemble. The key manipulation was correlation between angular sizes and distance to the observer, which changed perceived range (Tiurina & Utochkin, 2019). In Experiment 1 we used mirror stereoscopes to manipulate depth, while in Experiment 2 - Ponzo-like illusion.

2. Experiment 1

In Experiment 1, participants evaluated the size range of a briefly presented set of circles. To avoid different adaptation aftereffects to retinotopic and spatiotopic coordinates (Corbett & Melcher, 2014; Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Tiurina, Markov, Corbett, & Utochkin, 2019) both presented (sample) and tested (set with which participants evaluated range of the sample) sets were shown in depth with two types of size-distance correlation. Similarly to Tiurina & Utochkin, 2019, the main manipulation was the correlation between angular size and depth. In positive size-distance correlation small objects were presented closer to the observer and big objects were presented farther to the observer, and vice versa in negative size-distance correlation. If statistics were estimated, based on perceived sizes (Tiurina & Utochkin, 2019; Im & Chong 2009), then the perceived range increases with the positive correlation, because closely presented small objects appeared more smaller and far presented big objects appeared bigger, and decreases with negative correlation. The size-distance correlation of sample and test will differently change the evaluated range. For the sample: set with positive correlation will be perceived as having a broader range, so participants will overestimate their answer; while the set with negative correlation will be perceived as having a narrower range and there will be an underestimation. However, participants will respond using a test, which could also be presented in two different size-distance correlations. Thus, participants will perceive a test with a broader range (positive correlation) and will try to make it narrower (underestimation) to match their representation of the range of the sample, and will perceive a test ensemble with narrower range (negative correlation), which will try to make it broader (overestimation). So, we expect to find the maximum overestimation in condition with positive sample size-distance correlation and negative test size-distance correlation and the maximum underestimation in condition with negative sample size-distance correlation and positive test size-distance correlation.

2.1. Method

2.1.1. Participants

Twenty-one undergraduate students of the Higher School of Economics (18 female, the average age is 19.9 years) participated in the experiment for extra credits in a psychology course. The sample size was determined based on the sample sizes used in other studies, which focused on estimating range of the set (examples from different laboratories: (Dakin & Watt, 1997; Haberman, Lee, & Whitney, 2015; Morgan et al., 2008; Norman et al., 2015; Solomon et al., 2011; Suárez-Pinilla et al., 2018). All students reported having normal or corrected-to-normal visual acuity, stereo vision, and no neurological problems. Each participant passed a short stereo vision test before the experiment.

2.1.2. Apparatus and stimuli

The stimulation was developed and presented via PsychoPy for Linux Ubuntu (Peirce, 2007; Peirce et al., 2019) on a standard VGA-monitor (screen diagonal 19 in., 85 Hz refresh rate, resolution 1024 × 768 pixels). Mirror stereoscopes (Adjustable Stereo Wide Viewer - SS-1) were used for obtaining the stereo effect. Ophthalmic chin rests were used to fix participants' heads, which provided a constant viewing distance during the entire experiment. Observers responded by pressing keys on a computer keyboard and rotating the mouse wheel.

For stimuli, we used stereoscopic pairs. On a black background, two grey square fields (8.4° x 8.4°, 7.3 × 7.3 cm) were presented on the left and right part of the monitor (with the distance between the center of each square to the center of the screen equal to 7.5° (6.5 cm)). The black fixation point appeared in the center of each grey field.

A set of eight white circles for both sample and test were located on an imaginary circumference with a radius of 2° (1.7 cm). Each circle was centered at one of eight rotational positions on the circumference with jitter (−0.06° to 0.06°, −0.05 to 0.05 cm), starting at 0° and following one after another with a step of 45°. The center of the imaginary circumference was placed at the fixation point. The mean size of the circles was fixed across all trials and equal to 0.6° (0.05 cm). The size distribution consisted of eight sizes equally spaced along with Teghtsoonian's (1965) size scale. The range was calculated as a distance between the biggest and the smallest sizes in the set in the units of the mean. For the sample set, the range varied from 0.4 to 1.4 (with a step equal to 0.1) and for the test set - from 0 to 1.8 (with a step equal to 0.01). Thus, the biggest and smallest circles were ± ½ of the range away from the mean, the second biggest/smallest were ± $^3/_8$ of the range away from the mean, third - ± ¼ away of the range from the mean and middle sizes were ± $^1/_8$ of the range away from the mean.

We changed the depth of the stimuli by manipulating binocular disparity - different horizontal shifts of left-eye and right-eye images relative to the fixation cross. For the positive correlation condition, bigger sizes appear at the farthest plane and smaller sizes appear at the closest plane. So, for two smallest circles, the disparity (Disparity = Left shift — Right shift) was about −0.24° (−0.2 cm), making them appear closer than the plane of fixation, for the second two circles by size rank, the disparity was equal to −0.12°(−0.1 cm). For the third and fourth pairs of circles, disparities were equal to +0.12°(0.1 cm) and + 0.24° (0.2 cm), respectively, making them appear farther than the plane of fixation. For negative correlation condition, bigger sizes appear at the closest plane and smaller sizes appear at the farthest plane. Thus, the two biggest circles were assigned the −0.24°(−0.2 cm) disparity, etc.

2.1.3. Procedure

Experimental sessions were run in the darkened room. The participants seated at approximately 50 cm (19.7 in.) from the monitor. Each participant passed the binocular vision test using the same principle as Julesz's stereograms (Julesz, 1971). Before the beginning of the experiment, a mirror stereoscope was fixated with belts on a participant's head. Random sets of colored circles (blue, red and white) on a grey background were presented stereoscopically. Participants had to identify and name the color of circles standing out from the background due to the disparity. Success on this test served as a criterion for an ability to segment multiple items by the depth and correct stereoscope setup. Before the main experiment, there were 8 practicing trials.

The fixation cross was presented during all trials. In each trial, the sample was presented for 500 ms. 200 ms later, the test set was presented (See Fig. 1). Participants were asked to adjust the range by rotating a mouse wheel that increased or decreased the range of the test distribution with a step of 0.01. Mouse rotation changed the size of each circle in the set according to the principle described previously. Participants were instructed to press the SPACE button to confirm their response and finish a trial. The following trial started with pressing the SPACE button, so participants could progress at a comfortable pace and take a rest whenever they wanted. The order of trials of different conditions were randomly mixed to additionally avoid various adaptation aftereffects.

2.1.4. Design and data analysis

Experiment 1 had a 2 (sample size-distance correlation: positive vs. negative) × 2 (test size-distance correlation: positive vs. negative) within-subject design. 77 trials were presented for each condition. Thus, the total number of trials was equal to 308 per observer.

We estimated in each trial the *Bias* - a signed deviation between the estimated and correct range: $Bias = (Response - Correct)/Correct$, where the *Response* is the range of test ensemble set by a participant, and *Correct* is the actual range of angular sizes of the sample set. Thus, *bias* equal or close to zero means that there is no bias in the participant's response. While positive or negative bias indicates overestimation or underestimation of the range, respectively. We also calculated the precision of estimated range: $Precision = |Bias - M_{ij}(Bias)|$, where $M_{ij}(Bias)$ is an average bias in the i-th cell of experimental design for the j-th participant. Our predictions considered only changes in bias, while the precision of range estimation according to our predictions should remain stable across all conditions.

We applied the standard frequentist and Bayesian repeated measures ANOVA. The Bayesian approach estimates the odds of H1 to H0 (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The Bayes factor (BF_{10}) was calculated using JASP 0.9.0.0 (JASP Team, 2020; Wagenmakers et al., 2017) and interpreted using the standard Jeffrey's scale (1961). The Cauchy distribution with a width of 0.707 was used as a prior distribution of effect sizes under H0. A Bonferroni correction was made for multiple comparisons in calculating the statistical significance level.

2.2. Results

2.2.1. Bias

We found effect of sample size-distance correlation ($F(1,18)$ =

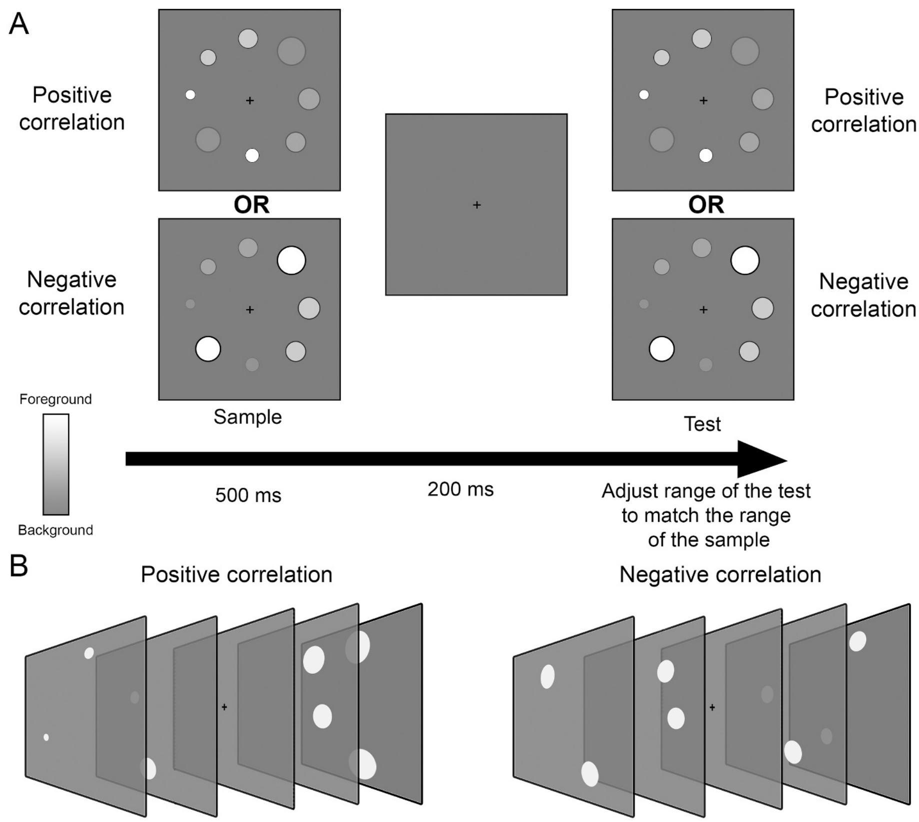

Fig. 1. (A) The time course of Experiment 1. Different contrasts of circles depict different distances. Black circle outlines are used for better image perception for the reader. (B) The distribution of circles over apparent distances with perceived sizes.

14.018, $p < .001$, $\eta^2_p = 0.009$, $BF_{10} = 34.08$, Fig. 2), test size-distance correlation ($F(1,18) = 4.061$, $p = .044$, $\eta^2_p = 0.003$, $BF_{10} = 0.231$) and also their interaction ($F(1,18) = 4.659$, $p = .031$, $\eta^2_p = 0.003$, $BF_{10} = 0.325$). Range of ensembles presented in positive correlation (tested in positive correlation: $M = 0.039$; tested in negative correlation: $M = 0.039$) were overestimated compared to negative correlation (tested in positive correlation: $M = -0.019$; tested in negative correlation: $M = 0.022$; comparison: $t(20) = 3.744$, $p < .001$, Bonferroni corrected $\alpha = 0.025$, Cohen's $d = 0.093$). While the range of ensembles tested in positive correlation were underestimated compared to negative

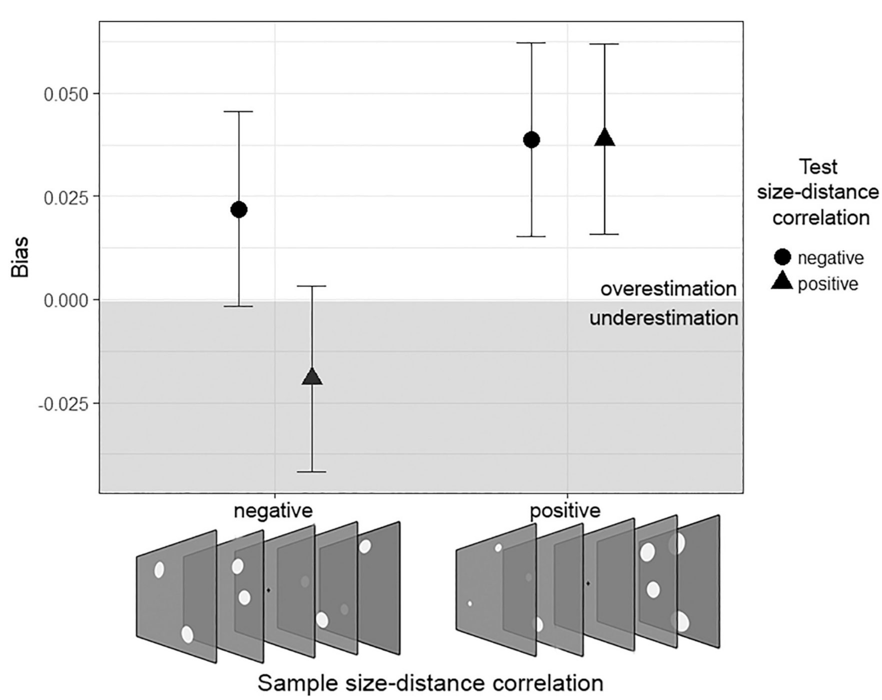

Fig. 2. An effect of the sample size-distance correlation and test size-distance correlation on bias in Experiment 1. Error bars denote 95% confidence intervals.

correlation ($t(20) = 2.015$, $p = .022$, Bonferroni corrected $\alpha = 0.025$, Cohen's $d = 0.05$).

As shown in Fig. 2, the highest underestimation was in the condition where the ensemble was presented in a negative correlation and tested by the ensemble with a positive correlation ($M = -0.019$). The highest overestimation was in conditions where the sample was presented in a positive correlation ($M = 0.039$).

2.2.2. Precision of range estimation

We found no effect of sample size-distance correlation (positive sample correlation: test with positive correlation: $M = 0.294$, test with negative correlation: $M = 0.290$; negative sample correlation: test with positive correlation: $M = 0.285$, test with negative correlation: $M = 0.297$; comparison: $F(1,18) = 0.032$, $p = .859$, $\eta^2_p = 0.000$, $BF_{10} = 0.029$), test size-distance correlation ($F(1,18) = 0.512$, $p = .474$, $\eta^2_p = 0.000$, $BF_{10} = 0.035$) and their interaction ($F(1,18) = 1.782$, $p = .182$, $\eta^2_p = 0.001$, $BF_{10} = 0.000$).

2.3. Discussion

The most important result of this experiment: the largest underestimation of range corresponds to the set with a negative correlation between angular size and depth, and test with a positive correlation between angular size and depth. When an observer perceived a set with a negative size-distance correlation - larger objects appeared smaller, while smaller objects appeared larger. Thus, observers had to underestimate range (relative to the range of the angular sizes of the set). However, observers reported range using the test also presented in positive or negative correlation. The perceived range of the test appeared significantly broader in the condition with the test presented in positive size-distance correlation, so to compensate for the difference in the representations of sample presented in negative size-distance correlation and test presented in positive correlation participants were forced to considerably decrease the range of the test to match their representation of sample. This result suggested that rescaling occur before range estimation for both sample and test.

We found interaction between sample and test's size-distance correlation, because there was no overestimation effect in the condition with a positive size-distance correlation and test with a negative size-distance correlation. Tiurina and Utochkin (2019) similarly did not find overestimation effect in positive correlation condition. This result could be related to the ceiling effect, global overestimation of the range and small range of apparent distances used in experiment, allowing only to observe underestimation effect in negative correlation condition. In general, the largest overestimation was in condition with the positive sample size-distance correlation in the sample. These results are consistent with our hypotheses that summary statistics estimation occurs after rescaling.

3. Experiment 2

Thus, in Experiment 1 and Tiurina and Utochkin (2019) study results were not symmetrical for correlation conditions and in both studies we used stereoscopic presentation (binocular depth cue). We conducted Experiment 2 with the monocular cue of depth – the image of corridor conducted similarly to Ponzo illusion (Ponzo, 1912; similar to (Fang et al., 2008; Murray et al., 2006; Ni, Murray, & Horwitz, 2014)). According to size constancy, in the three-dimensional scene near objects appear smaller, while distant objects appear bigger. However, when there are no real physical distance changes (or distance estimation is incorrect) and depth induced by perceptive cues (Ponzo illusion), in order to maintain size constancy (Gregory, 1963) perceived sizes change in compensation. Stereoscopic presentation significantly limits the sizes of the presentation area and the choice of object's sizes. Thus, in this experiment, we were able to use more variable sizes and bigger presentation area.

3.1. Method

3.1.1. Participants

Twenty-two undergraduate students of the Higher School of Economics (20 female, the average age is 19.3 years) participated in the experiment for extra credits in a psychology course. All students reported having normal or corrected-to-normal visual acuity and no neurological problems.

3.1.2. Apparatus and stimuli

The stimulation was developed and presented via PsychoPy for Linux Ubuntu (Peirce, 2007; Peirce et al., 2019) on a standard VGA-monitor (screen diagonal 19 in., 85 Hz refresh rate, resolution 1024 × 768 pixels).

Stimuli were presented on a grey background or the image of the corridor (21.8° × 21.8°, 19.2 × 19.2 cm, Fig. 3). The image of the corridor was generated using 3DsMax and Adobe Photoshop.

Set of eight white circles were located on the invisible 8 × 8 grid (7.5° × 7.1°, 6.5 × 6.2 cm), the center of the grid was shifted down by 4.2° (3.6 cm). The width of each cell was equal to 1° (0.9 cm). To avoid overlapping, each circle was quasi-randomly located in each cell in such a way that in each column and row only one object was presented and also objects in the neighboring rows didn't occupy adjacent columns. Coordinates of circles were calculated in such a way that the bottom of circles was located on the lines on the background image of the corridor.

To avoid cues from the mean of the sample set, the mean of the sample was randomly chosen from 0.3° to 0.7° (0.26 to 0.6 cm). The mean size of the test set was constant and equal to 0.5° (0.4 cm). Similarly to Experiment 1, sizes were equally spaced along with Teghtsoonian's (1965) size scale. The range was calculated as a distance between the biggest and the smallest sizes in the set in the units of the mean. For the sample set, the range was varied from 0.4 to 1.4 (with a step equal to 0.1) and for the test set - from 0 to 1.8 (with a step equal to 0.01).

For positive correlation condition, bigger sizes appear at the farthest plane and smaller sizes appear at the closest plane (Fig. 3). Thus, smaller circles were located closer to the bottom of the screen, while bigger circles were located closer to the middle of the screen. For negative correlation condition, smaller sizes appear at the farthest plane and bigger sizes appear at the closest plane – smaller circles were located closer to the middle of the screen, bigger – to the bottom of the screen. For no correlation condition and test set, there was no correlation between sizes and position of objects (depth).

3.1.3. Procedure

Experimental sessions were run in a darkened room. The participants seated at approximately 50 cm (19.7 in.) from the monitor. The fixation cross was presented during all trials. In each trial, the sample was presented for 500 ms. 200 ms later, the test set was presented. The instruction was the same as in Experiment 1. Before the main experiment, there were 6 practicing trials. The experiment was conducted in two blocks: with illusion background and with a grey background presented. Half of the observers firstly participated in the block with a grey background and after that in the block with illusion background, another half of observers vice versa. This block design allowed us to control possible aftereffects caused by presentation of illusion image.

3.1.4. Design and data analysis

Experiment 2 had a 3 (sample size-distance correlation: positive vs. negative vs. no correlation) × 2 (the type of background: the image of corridor vs. no image) within-subject design. 77 trials were presented for each condition. Thus, the total number of trials was equal to 462 per observer. We estimated bias and precision as in Experiment 1. Data analysis was identical to Experiment 1. We also analyzed bias with correction to baseline by estimating the difference between condition with present and absent image of corridor. This analysis will allow us to

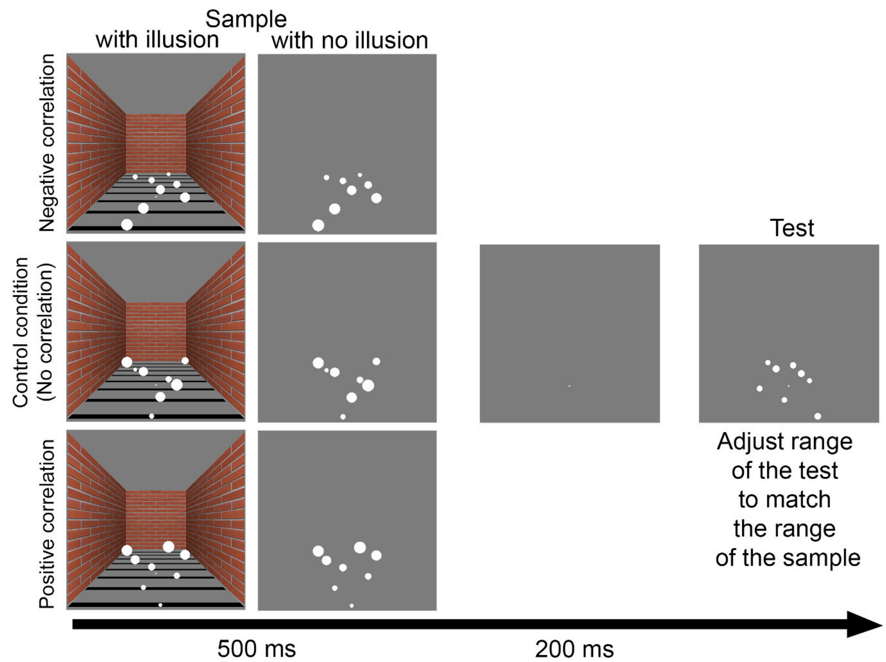

Fig. 3. The time course of Experiment 2. The figure depicts all six possible types of sample (3 types of size-distance correlation × 2 types of background: the image of corridor vs. no image).

compare three correlation conditions, because in Experiment 2, they are different not only by size-distance correlation, but also by circle's coordinates - in positive correlation bigger circles were closer to the center of the screen, while in negative correlation smaller circles were closer to the center of the screen. Thus, even with no illusion presented observers still had one depth cue - height, which could influence the results. Estimation of bias with correction to baseline will allow us control differences in coordinates between size-distance correlation conditions and investigate the influence of size-distance correlation on range estimation with control of differences in coordinates between these conditions.

3.2. Results

3.2.1. Bias

We found effect of size-distance correlation ($F(2,42) = 121.02$, $p < .001$, $\eta^2_p = 0.334$, $BF_{10} > 10^{40}$, Fig. 4) and also interaction of size-distance correlation and type of background ($F(2,42) = 20.683$, $p < .001$, $\eta^2_p = 0.079$, $BF_{10} > 10^5$). There was no significant effect of type of background ($F(1,21) = 0.259$, $p = .611$, $\eta^2_p = 0.001$, $BF_{10} = 0.065$). When there was no correlation (with illusion: $M = 0.108$; without illusion: $M = 0.108$) participants maximally overestimated range (compare

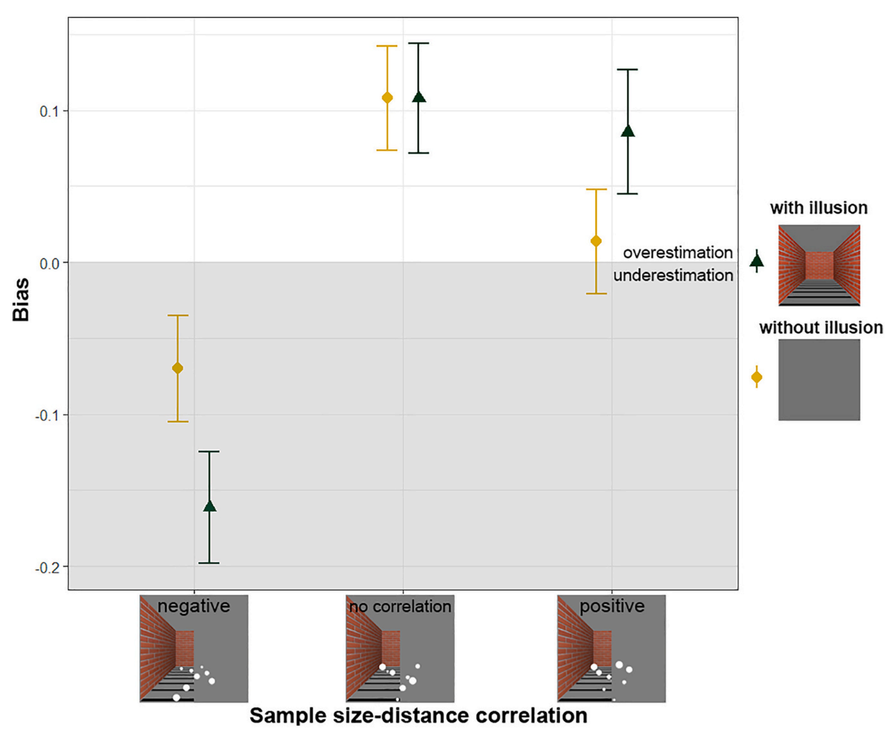

Fig. 4. An effect of the sample size-distance correlation and illusion on bias in Experiment 2. Error bars denote 95% confidence intervals.

to positive correlation: with illusion: $M = 0.086$; without illusion: $M = 0.014$; comparison: ($t(21) = 4.026$, $p < .001$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.259$, $BF_{10} = 362.1$); to negative correlation: with illusion: $M = -0.161$; without illusion: $M = -0.070$; comparison: ($t(21) = 14.613$, $p < .001$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.939$, $BF_{10} > 10^{40}$)). Participants underestimated range of ensemble in negative correlation compared to positive correlation ($t(21) = 11.076$, $p < .001$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.712$, $BF_{10} > 10^{20}$).

When size-distance correlation was negative, range was more underestimated in condition with presented illusion compared to background with no illusion ($t(21) = 5.125$, $p < .001$, Cohen's $d = 0.329$, $BF_{10} > 10^5$). When size-distance correlation was positive, range was more overestimated in condition with presented illusion compared to background with no illusion ($t(21) = 3.367$, $p < .001$, Cohen's $d = 0.216$, $BF_{10} = 17.2$). When only grey background was presented, range was more underestimated in condition with negative size-distance correlation in comparison to positive size-distance correlation ($t(21) = 4.423$, $p < .001$, Cohen's $d = 0.284$, $BF_{10} = 786.5$).

3.2.2. Precision of range estimation

We found no significant differences between positive (with illusion: $M = 0.189$; without illusion: $M = 0.169$), negative (with illusion: $M = 0.153$; without illusion: $M = 0.172$) and no correlation conditions (with illusion: $M = 0.179$; without illusion: $M = 0.173$; comparison: ($F(2,42) = 2.054$, $p = .129$, $\eta^2_p = 0.008$, $BF_{10} = 0.051$)), type of background ($F(1,21) = 0.080$, $p = .778$, $\eta^2_p = 0.000$, $BF_{10} = 0.073$) and their interaction ($F(2,42) = 2.701$, $p = .068$, $\eta^2_p = 0.011$, $BF_{10} = 0.17$).

To check whether the differences between size-distance correlations when the illusion was not presented was not caused by aftereffect – some observers firstly participated in the block with illusion, thus this could lead to some aftereffects in the second block – we additionally analyzed those participants ($N = 13$) who firstly take part in the block with a grey background. We found effect of the size-distance correlation ($F(2,24) = 81.957$, $p < .001$, $\eta^2_p = 0.366$, $BF_{10} > 10^{30}$) and also interaction of size-distance correlation and type of background ($F(2,24) = 11.325$, $p < .001$, $\eta^2_p = 0.074$, $BF_{10} = 99.2$). There was no significant effect of type of background ($F(1,12) = 0.156$, $p = .694$, $\eta^2_p = 0.001$, $BF_{10} = 0.08$). Thus, the pattern was similar to the main analyses and we also found that sets in the negative size-distance correlation was underestimated in comparison to the positive size-distance correlation when grey background was presented ($t(12) = 3.454$, $p < .001$, Cohen's $d = 0.289$, $BF_{10} = 25.68$).

3.2.3. Bias with baseline correction

We found the effect of size-distance correlation ($F(2,42) = 20.683$, $p < .001$, $\eta^2_p = 0.079$, $BF_{10} > 10^6$, Fig. 5). The maximum overestimation was found in condition with positive size-distance correlation ($M = 0.072$; compared to none correlation condition: $M = 0.000$; comparison: $t(21) = 2.826$, $p = .005$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.182$, $BF_{10} = 3.658$; to negative correlation condition: $M = -0.091$; comparison: $t(21) = 6.416$, $p < .001$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.412$, $BF_{10} > 10^6$). The maximum underestimation was found in condition with negative size-distance correlation (compared to none correlation condition: $t(21) = 3.591$, $p = .001$, Bonferroni corrected $\alpha = 0.017$, Cohen's $d = 0.231$, $BF_{10} = 90{,}118$).

Thus we found that in none correlation condition bias with baseline correction did not differ from 0 ($t(21) = 0.002$, $p = .999$, Cohen's $d = 0.00001$, $BF_{10} = 0.072$) and maximum overestimation and underestimation in positive and negative correlation conditions respectively.

3.3. Discussion

Our analysis of bias with baseline control demonstrated that illusion manipulation influenced perceived range as predicted - the maximum overestimation was found in positive size-distance correlation and maximum underestimation in negative size-distance correlation. In comparison to previous experiments (Tiurina & Utochkin, 2019) we found evidence that both (positive and negative) size-distance correlation manipulation affected bias, while none correlation condition caused bias equal to zero. Thus, this is a clear demonstration that range is estimated only after size-distance rescaling.

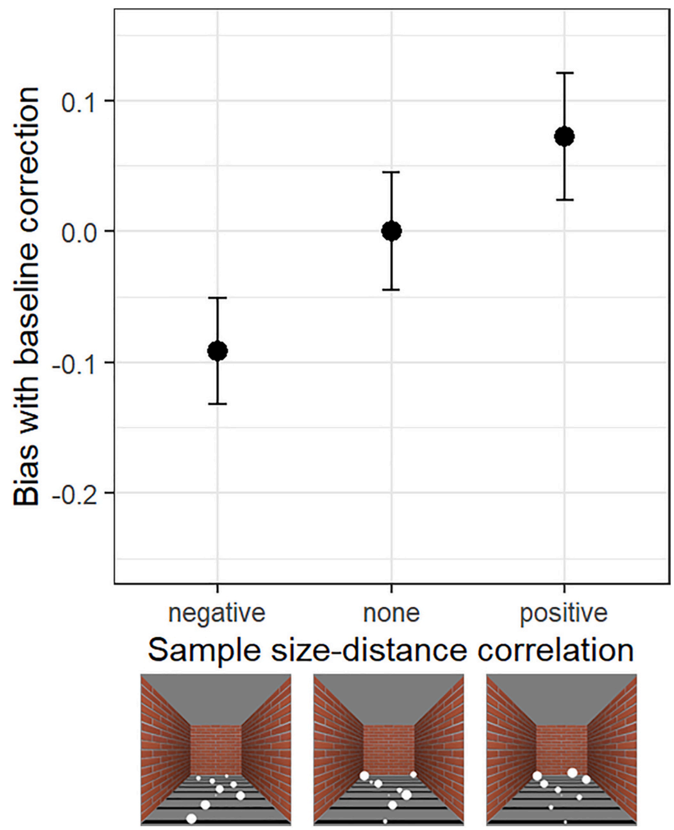

Fig. 5. An effect of the sample size-distance correlation in Experiment 2. Error bars denote 95% confidence intervals.

We also found that even without illusion present observers underestimated and overestimated range correspondingly to the negative and positive condition. One is possible explanation is that in positive and negative size-distance correlation location of circles was still correlated with size, thus this is one of the pictorial depth cues - height. The height as pictorial cue impacted the perceived range of the set, however, the effect from illusion manipulation was significantly higher (Bian, Braunstein, & Andersen, 2005; Epstein, 1966; Gardner, Austerweil, & Palmer, 2010). Additional analyses of participants who first took part in the block with a grey background demonstrated that height indeed was one of the depth cues and had an effect on the perception of range. This proves the automatic and inevitable character of the rescaling.

4. General discussion

Results of both experiments demonstrated that size-distance correlation manipulation affected the perceived range. In positive size-distance correlation condition perceived range increased: small objects, which were presented closer, appeared more smaller and big objects, which were presented farther, appeared more bigger, and in negative size-distance correlation perceived range decreased: small objects, which were presented farther, appeared more bigger and big objects vice versa. This change in size perception occurred through compensatory mechanisms of size constancy (Dwyer, Ashton, & Broerse, 1990; Gregory, 1963; Qian, Liu, & Lei, 2016; Qian & Petrov, 2016). Thus, we demonstrated that estimation of range of the set based on not "raw" retinal representations, but on already rescaled according to their distance's representations and that rescaling for multiple items occured at once. These results are consistent with previous findings (Tiurina & Utochkin, 2019; Im & Chong 2009) and demonstrate that ensemble summary statistics based on rescaled and detailed object's representations. Current study replicates and elaborates results of Tiurina and Utochkin (2019), using direct evaluation of the range instead of estimation of mean representation precision affected by changed apparent range. In Tiurina and Utochkin (2019) study, the effect was present only in one condition in both experiments and the effect size was small, while in the current study we demonstrated that both positive and negative size-distance correlation condition affected perceived range. We additionally controlled depth aftereffect in Experiment 1 (Corbett et al., 2012; Corbett & Melcher, 2014; Tiurina et al., 2019), cues from the mean (similar to Khvostov & Utochkin, 2019) and aftereffect caused by illusion by using blocked design in Experiment 2.

We observed higher bias and found that both size-distance correlation affected bias in experiment with monocular cues, compared to experiment with binocular cues. However, monocular cues and judgments (Marotta et al., 1997; McCann, Hayhoe, & Geisler, 2018; McKee & Taylor, 2010; Sperandio et al., 2012) are less powerful than binocular. Our stereoscopic presentation did not allow us to use a broader range of depth distances (and also depth perception using stereoscope could be incorrect for some observers (McKee & Taylor, 2010)), thus it could explain smaller bias in Experiment 1. Our results suggest that monocular cues (Ponzo illusion and height) could be used for future investigation of ensemble perception in depth, which is a significantly easier method and allows to study it using fMRI (Sperandio & Chouinard, 2015).

Despite height depth cue in Experiment 2, which similarly to Ponzo-like illusion induced the depth but was controlled by estimating bias with correction to baseline, there was another possible depth cue. Big white circles had higher contrast against grey background than small circle, thus it is possible that based on contrast differences big objects could appear closer than small objects - depth cue induced by brightness or contrast is called aerial perspective (Dresp-Langley & Reeves, 2012; O'Shea, Blackburn, & Ono, 1994; Qian & Petrov, 2012; Qian, Zhang, Wang, Li, & Lei, 2018; Rohaly & Wilson, 1999). However, if objects were rescaled according only to aerial perspective, we should observe the same results in positive and negative size-distance correlation conditions, because in both conditions big objects would always be perceived as closer and small objects as farther, causing decrease of perceived range. We do not rule out the possibility of some small influence of aerial perspective cue on our results, it could be a powerful pictorial cue but without other depth cues (O'Shea et al., 1994), thus, the main influence on results in Experiment 1 was from binocular disparity and in Experiment 2 - from Ponzo illusion and height.

We demonstrated that each angular size was rescaled according to individual depths before ensemble summary statistics estimation and apparently this process does not require focused attention. According to feature integration theory (Treisman, 1996) the binding of two visual dimensions requires attention. However, in contrast to feature integration theory predictions, recent studies of visual search proposed that search for conjunction of depth and other dimensions is parallel and automatic: apparent size (Proulx & Green, 2011) and depth (Aks & Enns, 1996; Nakayama & Silverman, 1986; Viswanathan & Mingolla, 2002) could capture and guide attention and possibly process preattentively (Wolfe, 1994; Wolfe & Horowitz, 2017; Wolfe & Utochkin, 2019). Interpretations of angular size and depth rely on each other in contrast to independent features, such as color and orientation. Thus, it is important to note that we cannot generalize our results to other visual dimensions and types of binding proposed in feature integration theory.

Size-distance rescaling possibly takes place in V1, where attentional modulation is highly limited (Luck, Chelazzi, Hillyard, & Desimone, 1997; Martínez et al., 1999; Slotnick, 2017), but still in size-distance rescaling feedback processing from higher brain areas requires. Fang et al. (2008) showed that attention could play an important role in the connection between high-level visual areas (LOC and PPA) and low-level visual areas (V1). Attention modulates the feedback from high-level visual areas, thus affecting size-distance rescaling in V1. In the current study, we do not directly investigate the role of attention in this process, but it seems that even under distributed attention top-down feedback still influences activation of size tuned neurons in V1. Focused attention influences the perceived depth and perceived size (Anton-Erxleben, Henrich, & Treue, 2007; Guan & Qian, 2020), so it is possible that focused attention just strengthens the link between size and depth, causing rescaling. Size-distance rescaling and summary statistics estimation does not require focused attention and occur under distributed attention (Baek & Chong, 2020; Robitaille & Harris, 2011; Chong & Evans, 2011; Chong & Treisman, 2005; Treisman, 2006). Ensemble summary statistics could be estimated on various levels of the visual system (Corbett & Melcher, 2014), thus it is possible that items could also interact with scene representation on later stages of the visual system - reducing the load on attention.

It is important to note that in the current study we do not investigate the representation of the mean, but the range. Previous studies have shown that the mean and range could be processed and estimated in parallel (Khvostov & Utochkin, 2019; Yang et al., 2018). However, there are recent studies (Jeong & Chong, 2020; Tong, Ji, Chen, & Fu, 2015), which demonstrated that representations of mean and range (variance) interact with each other. This contradiction in results between studies could be explained by that mean and range share similar basic mechanisms and their representations interaction occur after their independent computation (Jeong & Chong, 2020; Kim & Chong, 2020). This is consistent with neural models and recent studies (Haberman & Whitney, 2012; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018; Khayat & Hochstein, 2019; Kim & Chong, 2020; Utochkin, 2019), explaining mechanisms of ensemble perception by pooling and population encoding. Taken together with the previous findings (Tiurina & Utochkin, 2019; Im & Chong 2009), we can conclude that the range estimation, like the mean estimation, occurs after the rescale based on both binocular and monocular cues. According to our results, ensemble summary statistics studies (Haberman & Whitney, 2012; Hochstein et al., 2018; Khayat & Hochstein, 2019; Kim & Chong, 2020; Utochkin, 2019) and size constancy studies (Qian & Yazdanbakhsh, 2015; Sperandio & Chouinard, 2015) we could assume that neural mechanism underlying ensemble perception in depth is following: firstly retinal sizes in low visual areas (V1) are rescaled via feedback from higher brain areas (binocular cues: MT, LIP; monocular cues: LOC, PPA); secondly the feature signals from lower areas (V1) are pooled in higher region with large receptive fields, where information about distribution activation becomes explicit and available (Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015). Thus, our visual system does not represent single mean or range value (Kim & Chong, 2020), but rather represents population summary - distribution of neuron activation of higher brain areas with large receptive fields - with greater activation near the mean of the perceived set and with level of noise depending on the range of the perceived set. Thus, it is possible that other kinds of ensemble summary statistics processed similarly and computed based on perceived sizes.

Our study clearly demonstrates that ensemble summary statistics could be based on integrated objects representations and partly explains how ensemble summary statistics allows us to perceive detailed world representation with minimum attentional load and without individual objects processing (Alvarez, 2011; Cohen et al., 2016).

Open science statement

The data from all experiments reported in this article can be accessed at: https://osf.io/rgwxq/

CRediT authorship contribution statement

N.A.T. Conceptualization, Methodology, Software, Writing - Original

Draft, Writing - Review & Editing, Supervision, Funding acquisition. Y.A.M. Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Acknowledgements

The reported study was funded by RFBR according to the research project #18-313-00253. The authors thank Igor S. Utochkin for his helpful comments and mentorship. The authors thank Sang Chul Chong and anonymous reviewer for their helpful comments on earlier drafts of the article.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1467–1481. https://doi.org/10.1037/0096-1523.22.6.1467

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

Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision, 7*(11), 1–9. https://doi.org/10.1167/7.11.5

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception & Psychophysics, 77*(4), 1116–1131. https://doi.org/10.3758/s13414-015-0870-0

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86*(4), 2054–2068. https://doi.org/10.1152/jn.2001.86.4.2054

Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics, 82*(1), 63–79. https://doi.org/10.3758/s13414-019-01827-z

Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record, 59*(2), 171–186. Retrieved from http://opensiuc.lib.siu.edu/tpr/vol59/iss2/2/%5Cnpapers2://publication/uuid/5CB02E77-1797-45CD-8557-A91D8E46328A.

Bian, Z., Braunstein, M. L., & Andersen, G. J. (2005). The ground dominance effect in the perception of 3-D layout. *Perception & Psychophysics, 67*(5), 802–815. https://doi.org/10.3758/BF03193534

Boring, E. G. (1940). Size constancy and Emmert's law. *The American Journal of Psychology, 53*(2), 293. https://doi.org/10.2307/1417427

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

Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex, 25*(11), 4226–4239. https://doi.org/10.1093/cercor/bhu145

Cant, J. S., & Xu, Y. (2017). The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. *Journal of Cognitive Neuroscience, 29*(2), 398–412. https://doi.org/10.1162/jocn_a_01050

Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate). *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 634–638. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/wcs.136/full.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). 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

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences, 20*(5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006

Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics, 76*(3), 746–758. https://doi.org/10.3758/s13414-013-0595-x

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

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences, 24*(4), 87–186. 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

Deangelis, G. C. (2000). Seeing in three dimensions: The neurophysiology of stereopsis. *Trends in Cognitive Sciences, 4*(3), 80–90. https://doi.org/10.1016/S1364-6613(99)01443-6

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

Dresp-Langley, B., & Reeves, A. (2012). Simultaneous brightness and apparent depth from true colors on grey: Chevreul revisited. *Seeing and Perceiving, 25*(6), 597–618. https://doi.org/10.1163/18784763-00002401

Dwyer, J., Ashton, R., & Broerse, J. (1990). Emmert's law in the Ames room. *Perception, 19*(1), 35–41. https://doi.org/10.1068/p190035

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70*(6), 946–954. https://doi.org/10.3758/PP.70.6.946

Epstein, W. (1966). Perceived depth as a function of relative height under three background conditions. *Journal of Experimental Psychology, 72*(3), 335–338.

Epstein, W., Park, J., & Casey, A. (1961). The current status of the size-distance hypotheses. *Psychological Bulletin, 58*(6), 491–514. https://doi.org/10.1037/h0042260

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology, 18*(21), 1707–1712. https://doi.org/10.1016/j.cub.2008.09.025

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

Gardner, J. S., Austerweil, J. L., & Palmer, S. E. (2010). Vertical position as a cue to pictorial depth: Height in the picture plane versus distance to the horizon. *Attention, Perception, & Psychophysics, 72*(2), 445–453. https://doi.org/10.3758/APP.72.2.445

Gnadt, J. W., & Mays, L. E. (1995). Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three-dimensional space. *Journal of Neurophysiology, 73*(1), 280–297. https://doi.org/10.1152/jn.1995.73.1.280

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*,15(1), 20–25. doi:https://doi.org/10.1016/0166-2236(92)90344-8.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature, 199*(4894), 678–680. https://doi.org/10.1038/199678a0

Guan, W., & Qian, J. (2020). Exogenous spatial attention shortens perceived depth. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01724-9

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15*(4), 16. https://doi.org/10.1167/15.4.16

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

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In *From perception to consciousness: Searching with Anne. Treisman* (pp. 339–349).

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*(7), 572–576. Retrieved from http://pss.sagepub.com/content/17/7/572.short.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision, 15*(4). https://doi.org/10.1167/15.4.7

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2018). Comparing set summary statistics and outlier pop out in vision. *Journal of Vision, 18*(13), 12. https://doi.org/10.1167/18.13.12

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology, 54*(1), 21–37. https://doi.org/10.2307/1417790

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

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics, 71*, 375–384.

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

Im, H. Y., Tiurina, N. A., & Utochkin, I. S. (2020). An explicit investigation of the roles that feature distributions play in rapid visual categorization. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02046-7

JASP Team (2020). JASP (version 0.9) [computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research, 167*, 46–53.

Julesz, B. (1971). *Foundations of cyclopean perception*. Oxford: *University of Chicago Press*. University of Chicago Press. https://doi.org/10.1097/OPX.0b013e318155ab62

Kaufman, L., Kaufman, J. H., Noble, R., Edlund, S., Bai, S., & King, T. (2006). Perceptual distance and the constancy of size and stereoscopic depth. *Spatial Vision, 19*(5), 439–457. https://doi.org/10.1163/156856806778457377

Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception, & Psychophysics, 81*(8), 2850–2872. https://doi.org/10.3758/s13414-019-01792-7

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

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychological Review, 60*(4), 223–231. https://doi.org/10.1037/h0060882

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. https://doi.org/10.1037/xhp0000804

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception : A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision, 14*(8), 1–13. https://doi.org/10.1167/14.8.26.doi

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*(1), 24–42. https://doi.org/10.1152/jn.1997.77.1.24

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

Mack, A., & Rock, I. (1998). Inattentional blindness. *MIT Press/Bradford books series in cognitive psychology*, 5(3), Inattentional blindness. xiv, 273. doi:https://doi.org/10.1016/j.aorn.2010.03.011.

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142*(2), 245–250. https://doi.org/10.1016/j.actpsy.2012.11.002

Marotta, J. J., Behrmann, M., & Goodale, M. A. (1997). The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. *Experimental Brain Research, 116*(1), 113–121. https://doi.org/10.1007/PL00005731

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., … Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience, 2*(4), 364–369. https://doi.org/10.1038/7274

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

McCann, B. C., Hayhoe, M. M., & Geisler, W. S. (2018). Contributions of monocular and binocular cues to distance discrimination in natural scenes. *Journal of Vision, 18*(4), 12. https://doi.org/10.1167/18.4.12

McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision, 10*(10), 5. https://doi.org/10.1167/10.10.5

Mikellidou, K., Gouws, A. D., Clawson, H., Thompson, P., Morland, A. B., & Keefe, B. D. (2016). An orientation dependent size illusion is underpinned by processing in the extrastriate visual area, LO1. *I-Perception, 7*(5), 1–21. https://doi.org/10.1177/2041669516667628

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*(2), 343–352. https://doi.org/10.1037/h0043158

Mishkin, M., & Ungerleider, L. G. (1982). Two cortical visual systems. *Analysis of Visual Behavior*, 549–586.

Mon-Williams, M., Tresilian, J. R., McIntosh, R. D., & Milner, D. A. (2001). Monocular and binocular distance cues: Insights from visual form agnosia I (of III). *Experimental Brain Research, 139*(2), 127–136. https://doi.org/10.1007/s002210000657

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

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9*(3), 429–434. https://doi.org/10.1038/nn1641

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*(6059), 264–265. https://doi.org/10.1038/320264a0

Ni, A. M., Murray, S. O., & Horwitz, G. D. (2014). Object-centered shifts of receptive field positions in monkey primary visual cortex. *Current Biology, 24*(14), 1653–1658. https://doi.org/10.1016/j.cub.2014.06.003

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

O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research, 34*(12), 1595–1604. https://doi.org/10.1016/0042-6989(94)90116-3

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

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

Ponzo, M. (1912). Rapports entre quelques illusions visuelles de contraste angulaire et l'appréciation de grandeur des astres à l'horizon. *Archives Italiennes de Biologie, 58*, 327–329. https://doi.org/10.1016/j.neuroimage.2009.08.024

Pooresmaeili, A., Arrighi, R., Biagi, L., & Morrone, M. C. (2013). Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *Journal of Neuroscience, 33*(40), 15999–16008. https://doi.org/10.1523/JNEUROSCI.1770-13.2013

Proulx, M. J., & Green, M. (2011). Does apparent size capture attention in visual search? Evidence from the Muller-Lyer illusion. *Journal of Vision, 11*(13), 21. https://doi.org/10.1167/11.13.21

Puerta, A. M. (1989). The power of shadows: Shadow stereopsis. *Journal of the Optical Society of America A, 6*(2), 309. https://doi.org/10.1364/josaa.6.000309

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

Qian, J., Liu, S., & Lei, Q. (2016). Illusory distance modulates perceived size of afterimage despite the disappearance of depth cues. *PLoS One, 11*(7). https://doi.org/10.1371/journal.pone.0159228

Qian, J., & Petrov, Y. (2012). StarTrek illusion-general object constancy phenomenon? *Journal of Vision, 12*(2), 1–10. https://doi.org/10.1167/12.2.15

Qian, J., & Petrov, Y. (2016). A depth illusion supports the model of general object constancy: Size and depth constancies related by a same distance-scaling factor. *Vision Research, 129*, 77–86. https://doi.org/10.1016/j.visres.2016.09.015

Qian, J., & Yazdanbakhsh, A. (2015). A neural model of distance-dependent percept of object size constancy. *PLoS One, 10*(7), Article e0129377. https://doi.org/10.1371/journal.pone.0129377

Qian, J., Zhang, K., Wang, K., Li, J., & Lei, Q. (2018). Saturation and brightness modulate the effect of depth on visual working memory. *Journal of Vision, 18*(9), 16. https://doi.org/10.1167/18.9.16

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

Rohaly, A. M., & Wilson, H. R. (1999). The effects of contrast on perceived depth and depth discrimination. *Vision Research, 39*(1), 9–18. https://doi.org/10.1016/S0042-6989(98)00034-0

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

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. https://doi.org/10.3758/PBR.16.2.225

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059–1074. https://doi.org/10.1068/p281059

Slotnick, S. D. (2017). The experimental parameters that affect attentional modulation of the ERP C1 component. *Cognitive Neuroscience, 9*(1–2), 53–62. https://doi.org/10.1080/17588928.2017.1369021

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

Sperandio, I., & Chouinard, P. A. (2015). The mechanisms of size constancy. *Multisensory Research, 28*(3–4), 253–283. https://doi.org/10.1163/22134808-00002483

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience, 15*(4), 540–542. https://doi.org/10.1038/nn.3069

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision, 18*(7), 4. https://doi.org/10.1167/18.7.4

Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology, 78*(3), 392–402.

Tiurina, N. A., Markov, Y. A., Corbett, J. E., & Utochkin, I. S. (2019). The adaptation aftereffect of mean size precedes size-distance rescaling. *Perception, 48*, 128.

Tiurina, N. A., & Utochkin, I. S. (2019). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. *Journal of Experimental Psychology: General, 148*(4), 728–738. https://doi.org/10.1037/xge0000485

Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision, 15*(4). https://doi.org/10.1167/15.4.15

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*(2), 171–178. https://doi.org/10.1016/S0959-4388(96)80070-5

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

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

Trotter, Y., Celebrini, S., & Durand, J. B. (2004). Evidence for implication of primate area V1 in neural 3-D spatial localization processing. Journal of Physiology Paris, 98(1-3 SPEC. ISS.), 125–134. doi:https://doi.org/10.1016/j.jphysparis.2004.03.004.

Tsotsos, J. (1988). A "complexity level"analysis of immediate vision. *International Journal of Computer Vision, 1*, 303–320. https://doi.org/10.1007/BF00133569

Utochkin, I. S. (2019). The properties of large receptive fields can explain ensemble-related phenomena. *PERCEPTION, 48*, 15.

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., & 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), Article e0185452. https://doi.org/10.1371/journal.pone.0185452

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception, 31*(12), 1415–1437. https://doi.org/10.1068/p3432

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). *Bayesian inference for psychology* (pp. 1–19). Part II: Example applications with JASP. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-017-1323-7

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*(5), 931–941.

Weidner, R., Plewan, T., Chen, Q., Buchner, A., Weiss, P. H., & Fink, G. R. (2014). The moon illusion and size-distance scaling-evidence for shared neural patterns. *Journal of Cognitive Neuroscience, 26*(8), 1871–1882. https://doi.org/10.1162/jocn_a_00590
Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8*(6), 820–827. https://doi.org/10.1038/nn1461
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
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., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058
Wolfe, J. M., & Utochkin, I. S. (2019). What is a preattentive feature? *Current Opinion in Psychology, 29*, 19–26. https://doi.org/10.1016/j.copsyc.2018.11.005
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. doi:https://doi.org/10.1177/2041669517747297.