Running Head: ENSEMBLE PERCEPTION IN DEPTH

Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging

Natalia A. Tiurina and Igor S. Utochkin

National Research University Higher School of Economics, Russian Federation

Author Note

The reported study was funded by RFBR according to the research project #18-313-00253. The study was partially reported at the 16th Annual Vision Sciences Society. The abstract is published online:
Tiurina, N. A., & Utochkin, I. S. (2018). Size averaging is based on distal, not proximal object sizes. Journal of Vision, 16(12): 56. doi: 10.1167/16.12.56.

Abstract

Previous studies have shown that people are good at rapidly estimating ensemble summary statistics, such as the mean size of multiple objects. In the present study, we tested whether these average estimates are based on "raw" retinal representations (proximal sizes) or on how items should appear based on context, such as the viewing distance (distal sizes). In our experiments, observers adjusted the mean size of multiple objects presented at various apparent distances through a stereoscope. In Experiment 1, all items were shifted in depth by the same amount while the adjustable probe stayed at the fixed middle position. We found that presenting ensembles in an apparently remote plane made observers overestimate the mean size which is consistent with angular sizes being rescaled to distance. In Experiment 2, we presented individual sizes in different planes. While angular sizes and apparent distances were kept controlled across conditions, we only manipulated correlations between them. These manipulations affected the precision of size averaging in line with changes in the range of apparent rather than angular sizes. This pattern is possible only if the visual system rescales each individual size to its distance prior to averaging. Our finding demonstrates that ensemble summaries of basic features, such as size, can be based on quite elaborated representations of multiple objects. We also discuss important implications for size constancy.

Keywords: ensemble perception, ensemble summary statistics, size averaging, size constancy, rescaling, feature binding

Our visual system permanently deals with the huge amount of visual information about the world that requires very fluent processing of a lot of information at every moment of perception. Meanwhile, an ability to process the information deeply seems to be limited by the capacity of attention (Mack & Rock, 1998; Pylyshyn & Storm, 1988; Simons & Chabris, 1999) and working memory (Cowan, 2001; Luck & Vogel, 1997; Miller, 1956). The representation of numerous objects as ensemble summary statistics can be a tool to overcome these limitations (Alvarez, 2011) and provide an impression of seeing all these objects rather correctly at once (Cohen, Dennett, & Kanwisher, 2016). An ability to rapidly estimate an average feature of multiple objects, such as the mean size (Ariely, 2001; Chong & Treisman, 2003, etc), average brightness (Bauer, 2009), average orientation (Alvarez & Oliva, 2008; Attarha & Moore, 2015), average speed (Watamaniuk & Duchon, 1992), even average emotion or gender of a crowd of faces (Haberman & Whitney, 2007) is one of the most studied types of ensemble statistics. The average can be a better estimate of the collection of objects than any of those picked from individual objects (Alvarez, 2011; Ariely, 2001; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Ensemble averaging can take place rather early in perception (Chong & Treisman, 2003; Robitaille & Harris, 2011; Whiting & Oriet, 2011; but see Gorea, Belkoura, & Solomon, 2014). Beyond explicit tests of ensemble statistics, recent studies have shown their important role in other visual domains, such as the deployment of attention during visual search (Chetverikov, Campana, & Kristjánsson, 2017; Corbett & Melcher, 2014; Utochkin & Yurevich, 2016; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012) or encoding and storage in working memory (Brady & Alvarez, 2011; Corbett, 2016).

While the studies of ensemble perception have covered a broad range of basic sensory dimensions and more complex and abstract properties (such as facial features, Haberman & Whitney, 2007; or even animacy, Leib, Kosovicheva, & Whitney, 2016), most of them tested each dimension independently. Only a few published works were aimed to study how ensemble summaries are extracted from a set of objects varying along more than one dimension (Emmanouil & Treisman, 2008; Huang, 2015; Oriet & Brand, 2013; Utochkin & Vostrikov, 2017).

One important reason to consider ensemble perception beyond unidimensional feature distributions is that, in real-world perception, our visual system rarely deals with multiple items varying in only one basic dimension, even when one is attempting to estimate only one. For example, real objects are normally not shown on a flat screen but can be located at various distances from the observer, which makes the statistics of retinal sizes (as usually measured in the laboratory) not a veridical descriptor of actual size statistics. For more realistic estimates, the retinal sizes should be rescaled by the corresponding distances to obtain an impression of true sizes which are then summarized as an ensemble. An ability to rescale the attributes of retinal images (proximal stimuli) to get access to the attributes of the physical objects (distal stimuli) is referred to as *perceptual constancy*.

Size constancy is usually close to ideal under normal observation conditions (Holway & Boring, 1941; see Chouinard & Sperandio, 2015, for review) suggesting that our visual system is efficient in (1) estimating the distance based on the available pictorial and binocular cues (see also Kaufman et al., 2006) and in (2) linking between these cues and the retinal size. In line with this phenomenology of accurate rescaling, recent neurophysiological studies have shown evidence for the low-level of neural representation of size. Specifically, these studies found that V1 neurons tend to respond to the apparent (subjective equivalent of the physical) rather than retinal size of an object (Murray, Boyaci, & Kersten, 2006; Ni, Murray, & Horwitz, 2014; Sperandio, Chouinard, & Goodale, 2012). However, despite this seeming ease and the early character of size-distance rescaling, it represents a version of a computationally demanding task (Tsotsos, 1988) related to the "*binding problem"* (Treisman, 1996; Wolfe & Cave, 1999): While separate features of the visible world are efficiently processed by the independent modules of the visual system (Yantis, 2014), an ambiguity can arise regarding how to combine these separate feature representations in a coherent object representation. According to Treisman (1996), rescaling of a retinal size by its apparent distance is an example of *"conditional binding"*. Conditional binding implies that one

visible property (apparent distance in our case) influences the interpretation of another (that is, perceived size).

Viewing size rescaling as a case of the binding problem and given the existing debate about the mechanisms of binding, we may question the seemingly easy character of size rescaling. The core subject of this debate concerns the role of a limited-capacity mechanism, such as focused attention, in correct binding. Some theorists (e.g., Treisman, 1996, 2006; Treisman & Gelade, 1980; Wolfe & Cave, 1999; Wolfe, Võ, Evans, & Greene, 2011) suggest that focused attention is necessary for correct feature binding, localization, and recognition. Others (e.g., Di Lollo, 2012; Rosenholtz, Huang, & Ehinger, 2012) suggest that the binding problem can be solved by the visual system without appeal to an attentional mechanism (for example, via the activation of conjunction-selective neurons or via feedback links from coherent high-level perceptual "hypotheses", Di Lollo, 2012). Applying this to size rescaling, it was shown that attention indeed can participate in the conditional binding between size and depth cues. Fang, Boyaci, Kersten, and Murray (2008) demonstrated that the V1 size response is rescaled by distance information only when the test object is attended to; but when attention is diverted from that object, the V1 response is more consistent with the retinal size representation. This finding implies a possibility that size-depth binding is not completely automatic and attention-free.

Ensemble perception somewhat sharpens the problem of conditional binding in size rescaling. In typical size constancy experiments, participants are usually asked to compare between the size of a test stimulus with the size of a sample stimulus (see Chouinard & Sperandio, 2015, for review). That is, observers can serially compare between these stimuli giving full attention to each stimulus and its distance cues, binding them correctly. But in ensemble perception the observer can encounter a lot of objects at one time, each having its own angular size and its own distance. Whether the perception of mean angular size and other basic features has capacity limits is a debated question by its own (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2008; Attarha & Moore, 2015; Attarha, Moore, & Vecera,

2014; Chong, Joo, Emmmanouil, & Treisman, 2008; Jackson-Nielsen, Cohen, & Pitts, 2017; Marchant, Simons, & De Fockert, 2013; Maule & Franklin, 2015; Myczek & Simons, 2008; Simons & Myczek, 2008; Utochkin & Tiurina, 2014; Whitney & Leib, 2018), which becomes even more complicated when distance variation is added to size variation. Not only "knowing" all of the distances and angular sizes, but also "knowing" which distance goes with which size and to rescaling the size correspondingly is where binding can become especially challenging (Tsotsos, 1988; Treisman & Gelade, 1980; Treisman, 2006; Wolfe & Cave, 1999). From this point, our basic research questions follow. Can size-distance rescaling survive the capacity bottleneck, so that the sizes of multiple objects simultaneously are perceived in accordance with their apparent distances? Are ensemble summaries of size based on retinal or on rescaled sizes?

Im and Chong (2009) reported some evidence for size rescaling in ensemble perception. Their participants compared mean sizes of groups of circles surrounded by bigger or smaller irrelevant circles, typically inducing a contrast size illusion referred to as the Ebbinghaus illusion. Im and Chong (2009) found systematic biases in perceived mean sizes. The magnitudes of these biases were similar to those obtained when individual Ebbinghaus configurations were compared. This led Im and Chong (2009) to cautiously suggest that the mean size representation is computed from individually rescaled sizes. However, given the debated nature of rescaling under the Ebbinghaus illusion (Ehrenstein & Hamada, 1995; Jaeger, 1978; Jaeger, Klahs, & Newton, 2014; Roberts, Harris, & Yates, 2005; see Robinson, 1998, for review), it is difficult to say whether the results of Im and Chong (2009) are related to size constancy. In addition, as we will demonstrate below (see rationale of Experiment 2), Im and Chong's (2009) design does not sufficiently address the question whether individual sizes are actually rescaled prior to averaging.

To answer our questions about the capabilities of multiple size rescaling and ensemble perception based on this rescaling, we ran two experiments. In both experiments, observers had to estimate the average size of a set of differently sized circles presented at different apparent distances. We manipulated binocular distance depth cues using stereoscopic presentation (Julesz,

1971). In Experiment 1, we tested whether the average size is rescaled to match depth variations. In Experiment 2, we tested whether the overall ensemble rescaling is based on the local rescaling of individual sizes by individual distances.

Experiment 1

In Experiment 1, observers were briefly shown sets of differently sized circles and had to adjust the size of a probe item to match the mean size of the set. The sets could be presented at various apparent distances from the observers, while the probe was always at a fixed standard distance. This led to various depth relations between the sets and the probe which were critical for testing rescaling. If observers rescale the mean size given the distance, then the adjusted mean size would be biased to larger or smaller values depending on whether the sample set is presented farther or closer than the probe. Specifically, there should be a bias towards larger estimates if the sample set is presented farther than the probe, and vice versa. By contrast, if the mean size is calculated based on the raw retinal image and not rescaled then we would not expect any bias associated with the distance.

Method

Participants

Thirty undergraduate students of the Higher School of Economics (18 female, average age is 18.5 years) participated in the experiment for extra credits in a psychology course. The sample size was determined based on an upper medium range of sample sizes used in studies of visual averaging in different laboratories (examples: Attarha & Moore, 2015; Attarha et al., 2014; Brady & Alvarez, 2011; Brand, Oriet, & Sykes Tottenham, 2012; Chetverikov et al., 2016; Corbett, 2016; Corbett & Melcher, 2014; De Gardelle & Summerfield, 2011; Im & Halberda, 2013; Jackson-Nielsen et al., 2017; Robitaille & Harris, 2011; Utochkin & Tiurina, 2014; Yamanashi Leib et al., 2016, etc.). 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 prior to the experiment.

Apparatus and stimuli

The stimulation was developed and presented via PsychoPy v1.82 for Linux Ubuntu (Peirce, 2007) on a standard VGA-monitor (screen diagonal 19 inches, 85 Hz refresh rate, resolution 800 × 600 pixels). Mirror stereoscopes were used for obtaining the stereo effect. Ophthalmic chin and forehead rests were used to fixate participants' heads that provided a constant viewing distance during the entire experiment. Observers responded by pressing keys on a computer keyboard with their dominant hand.

Stereoscopic pairs were used as stimuli. The monitor with black background was separated horizontally in two halves relative to the center of the monitor screen. Each half contained a grey square field (14.57º × 14.57º), distances between the center of each such field and the center of the screen were 13º. The black fixation point appeared in the center of each grey field. A set of eight white circles with sizes ranging from .73º to 2.04º was used as a sample stimulus. The circles were located on an imaginary circumference with a radius of 3.6º. Each circle was centered at one of eight rotational positions on the circumference, starting at 0 degrees and following one after another with a step of 45 degrees. The center of the imaginary circumference was placed at the fixation point (except for horizontal shift providing binocular disparity between the eyes), so every circle was presented equidistantly from the fovea. The average angular size of the sample set varied between 1.1º and 1.8º. A single white circle with an adjustable size presented at fixation was used as a probe stimulus. The initial size of the probe was randomly set between .5º and 2.5º.

An apparent depth of the stimuli was achieved by manipulating binocular disparity, unequal horizontal shifts of right-eye and left-eye images relative to fixation. There were three possible depth conditions defined by the disparity of sample sets (Figure 1). In the *Background* condition, the imaginary circumference with all sample circles was shifted by –.15º on the left image and to .15º on the right image (negative and positive values indicate shift to the left and to the right,

respectively). This provided the total disparity of –.3º (*Disparity = Left shift – Right shift*) and led to the perception of the circles in a plane behind the fixation point. In the *Foreground* condition, disparity was +.3º. This provided the perception of the circles in a plane before the fixation point. The absolute disparity of .3º provides the satisfactory above-chance subjective "singleness" of disparate retinal images within the eccentricity range we used (Duwaer & Van Den Brink, 1981; Palmer, 1961). Finally, in the *Middle* condition, the images had a zero disparity (the circumferences were centered exactly on fixation), so the circles were perceived at the plane of fixation (Figure 1).

The right and left images of the probe circle always had a zero disparity, so it was always perceived as presented in the plane of fixation.

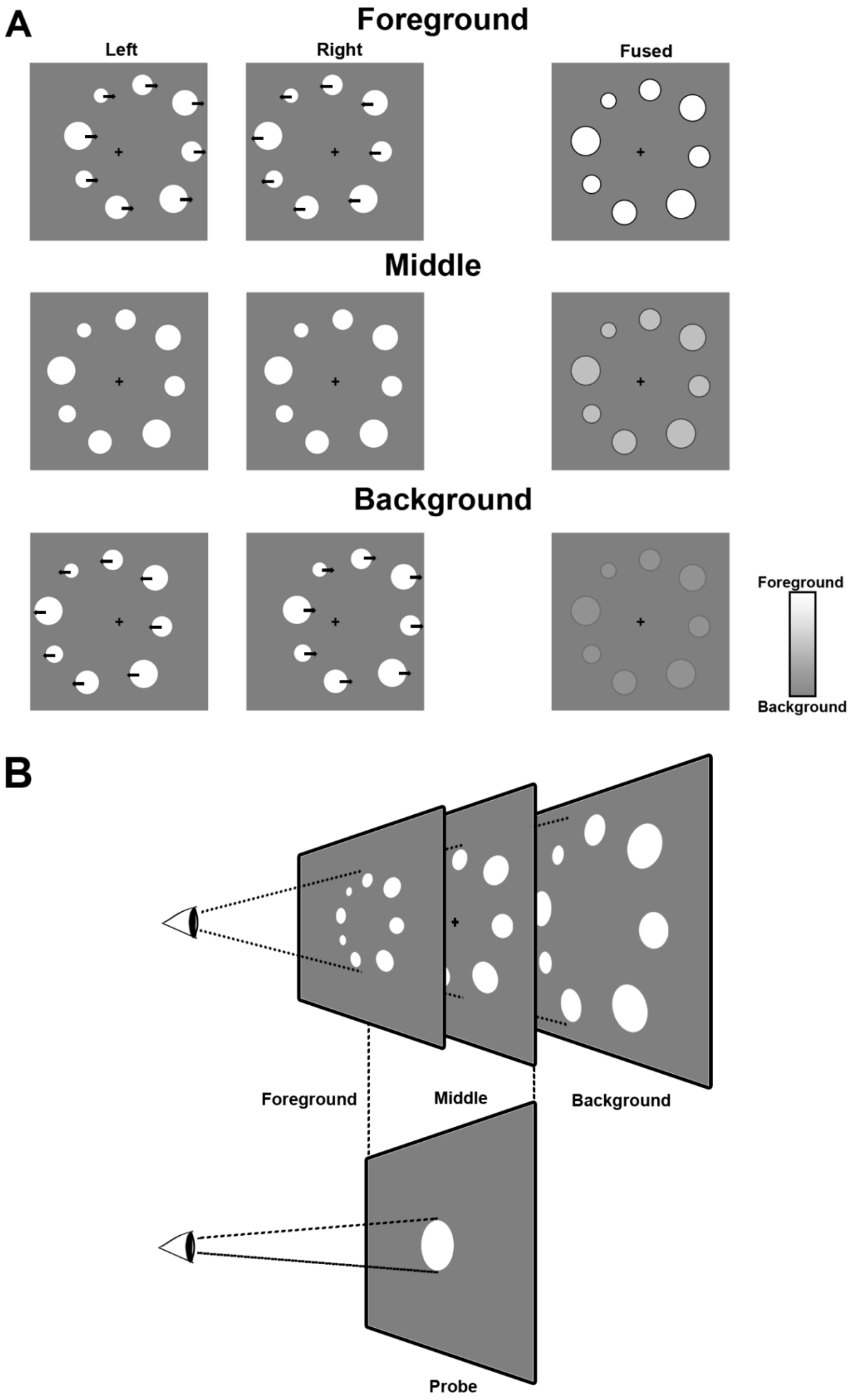
A
Foreground
Left
Right
Fused
Middle
Background
Foreground
Background
B
Foreground
Middle
Background
Probe

Figure 1. An example of stimuli used in Experiment 1. (A) Stereoscopic pairs for three depth conditions and corresponding fused frontal views of exactly the same set of circles as a function of binocular disparity. Arrows (not shown to observers) illustrate the directions of relative shifts of set members in a left- and a right-eye images. Different contrast levels on the fused images are used to show distance variations and black outlines are used to make low-contrast images more visible to the reader; in real stimulation, all circles had same contrast levels and were uniformly white. (B) Side views of sets at the three planes as they should be seen by a participant (assuming rescaling).

Procedure

Before the beginning of the experiment, a stereoscope was fixated with belts on a participant's head. The participants then seated at approximately 50 cm (19.7 inches) from the monitor placing the head on a chin rest. Each participant passed a binocular vision test using the same principle as Julesz's stereograms (Julesz, 1971). Random sets of white circles on a grey background were presented stereoscopically. Participants had to identify and name geometrical shapes standing out from the background due to disparity. Success on this test served as a criterion for an ability to segment multiple items by depth, which is exactly what was required for the experiment.

Each trial of the experiment started with the presentation of the fixation point that participants were instructed to look at. 500 ms later, a sample set was presented for 1000 ms followed by the presentation of a probe. The participants were asked to adjust the size of the probe so that it seemed equal to the mean size of all circles in the sample set. Adjustment was made by pressing "right arrow" or "left arrow" keys on a computer keyboard. Each press of a button increased or decreased the diameter of the probe by one pixel (~.05º). When the participants reached the apparent mean size, they were instructed to press SPACE to confirm their response and quit a trial. A next trial started upon pressing SPACE, so participants could progress in a comfortable pace and take a rest whenever they wanted. The experimental session consisted of 15 practicing trials followed by the experimental block of 150 trials.

Design and data analysis

Experiment 1 had a one-way within-subject design. The independent variable, Plane, included three conditions defined as depth planes of sample set (foreground, background, and middle). For

each condition, 50 trials were presented. In each trial, a signed deviation between the estimated and the correct mean sizes was calculated as a measure of response bias: *Bias* = (*Response* - *Correct*)/*Correct*, where *Respons*e is a probe size set by an observer, and *Correct* is the true mean of angular sizes of the sample circles. Here, *Bias* of about zero indicates an unbiased response, while positive or negative *Biases* indicate overestimation or underestimation, respectively. Changes in the magnitude and sign of *Bias* as a function of the depth plane were considered to be principal measures of rescaling. Assuming that the average bias reflects a systematic shift in the subjective perception of mean size, we calculated the precision of averaging as an absolute deviation (error) of a response from this equivalent in each trial: $Error = |Bias - M_{ij}(Bias)|$, where $M_{ij}(Bias)$ is an average *Bias* in the *i*-th cell of experimental design for the *j*-th participant.

To statistically estimate an effect of Plane on *Bias* and *Error* we used the standard frequentist and Bayesian ANOVA. The Bayesian ANOVA is a direct way to estimate evidence for H_1 against H_0 (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The Bayes factor (BF_{10}) is the odds between the relative likelihoods of H_1 and H_0 under the observed data, was calculated using JASP 0.8.2 (JASP Team, 2017; Wagenmakers et al., 2017). For Bayesian ANOVA, a prior was set at .5 on the *r* scale for fixed effects, as recommended by JASP developers (JASP Team, 2017; Wagenmakers et al., 2017). For pairwise *t*-tests within ANOVA, the Cauchy distribution with a width of .707 was used as a prior distribution of effect sizes under H_0 (JASP Team, 2017; Wagenmakers et al., 2017).

Results and discussion

We found a strong effect of Plane on *Bias* ($F(2,58) = 15.15$, $p < .001$, $\eta^2 = .343$, $BF_{10} = 2{,}35\times10^6$). As shown in Figure 2, it was provided by a greater positive (overestimation) bias in the Background condition compared to Foreground ($t(29) = 4.97$, $p < .001$, Bonferroni-Holm corrected $\alpha = .017$, Cohen's $d = .907$, $BF_{10} = 828$) and Middle ($t(29) = 4.15$, $p < .001$, Bonferroni-Holm corrected $\alpha = .025$, $d = .758$, $BF_{10} = 109$). We also found no evidence for an effect of the Plane on *Error* ($F(2,58) = .801$, $p = .454$, $\eta^2 = .027$, $BF_{10} = .007$).

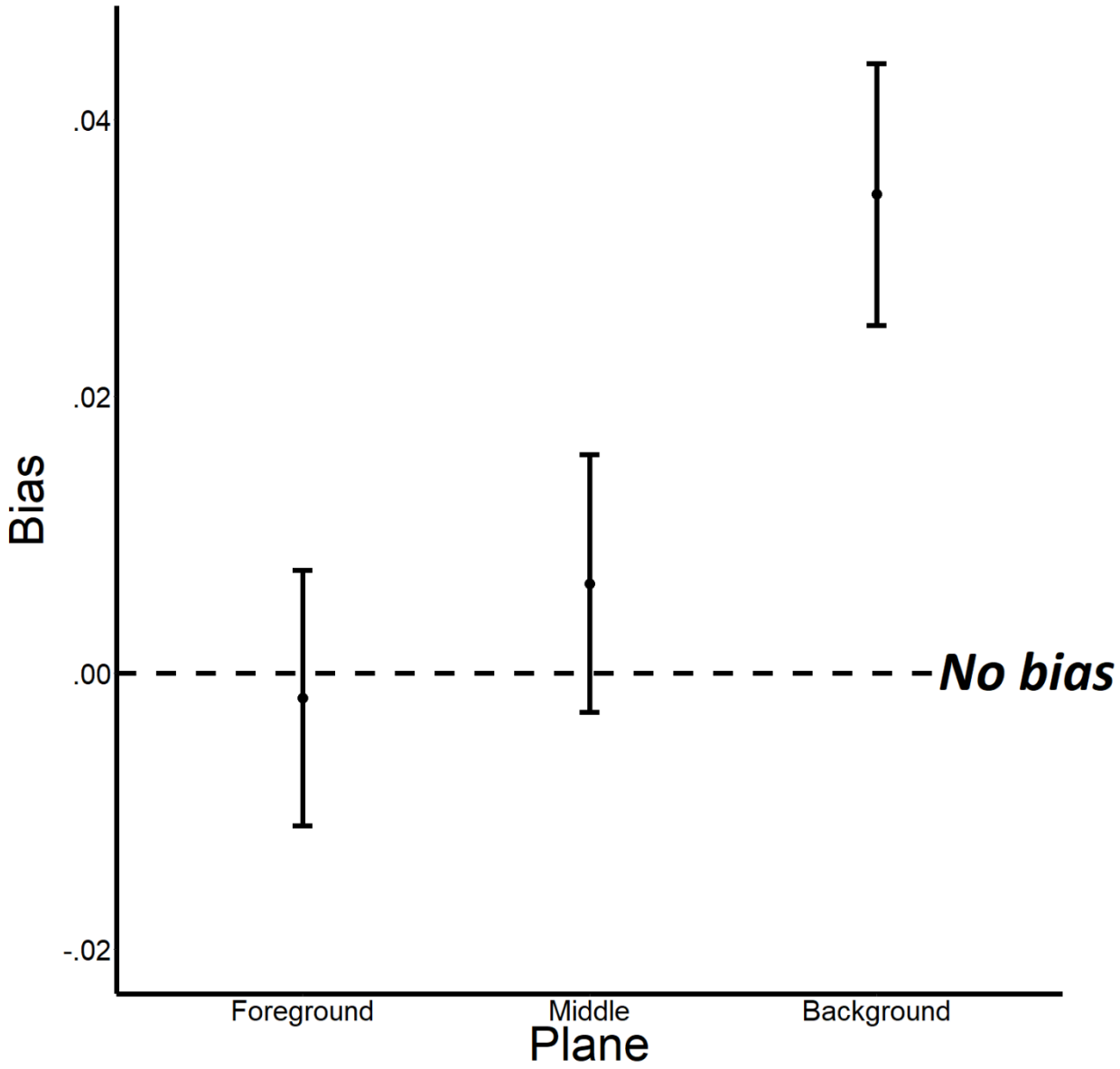

Figure 2. An effect of the apparent depth on *Bias* in Experiment 1. Error bars denote 95% CIs.

Our main finding in Experiment 1 is that presenting the set at the background elicited systematic overestimation of mean size, which is consistent with the prediction in accordance with the rescaling account. We failed to find a symmetric underestimation for sets at foreground compared to the middle plane. A possible explanation can be that isodisparate images, that we used to induce the perception of the foreground and the background, geometrically (according to the *angle bisector theorem*) correspond to non-symmetrical distances to both depth directions from the plane of fixation. Specifically, for a fixed disparity magnitude, the apparent distance between the foreground and the middle plane is shorter than the distance between the background and the

middle plane. That implies that angular sizes should be rescaled by factors corresponding to the apparent depth distances from the middle plane. Consequently, background items are rescaled to a greater extent than foreground items. Given the noise originating from the discrimination of depth and individual sizes (Kaufman et al., 2006; Mckee & Welch, 1992), as well as from size averaging (Alvarez, 2011; Im & Halberda, 2013), the differences between rescaled mean sizes in foreground and the middle planes could be insufficient to reveal a systematic bias, whereas the differences between rescaled mean sizes in background and the middle planes was sufficient to reveal such a systematic bias. Although the observed bias suggests that participants do somehow rescale their retinal images of multiple items when calculating the mean size, does this mean that the visual system operates on multiple rescaled representations when calculating ensemble statistics? That is, are individual retinal sizes bound with their distances prior to entering the averaging stage? Experiment 1 does not lead to an unambiguous answer to these questions. As we manipulated the depth of the entire set rather than the depths of individual items, we cannot estimate to which degree the individual depth information is taken into account. It is possible that observers in fact estimate the mean retinal size and rescale it to a single distance that is shared by all items. This possibility is in line with constant averaging precision (*Error*) that we observed in Experiment 1. Alternatively, all items are bound (rescaled) to their individual distances and the rescaled outputs are then averaged. The similar ambiguity concerns the earlier demonstration of mean size rescaling under the Ebbinghaus illusion (Im and Chong, 2009): If all to be averaged items are surrounded by inducers of same size (all inducers are either small, or large within a set), it is unclear whether the visual system rescales each individual size prior to averaging or computes the “retinal” mean and then rescales it to the common inducer context. A successful experimental dissociation between these two possibilities should involve independent manipulations, whereby any physical size can be associated with any distance within the same group. We implemented this approach in Experiment 2.

Experiment 2

An obvious way to test whether the visual system rescales each individual size to its individual depth prior to averaging, is presenting items in different planes and testing whether this would cause biases in mean size estimates. However, it seems to be a tricky manipulation. To illustrate this, consider a simplified example: An observer is presented with three circles with sizes $S_1 = 1$, $S_2 = 2$, $S_3 = 3$, that can be presented at three possible distances $D_1 = 1$, $D_2 = 2$, $D_3 = 3$. The mean retinal size is always $S_M = 2$. If we place S_1 at D_1, S_2 at D_2, and S_3 at D_3, the apparent sizes will become 1, 4, and 9, respectively, so the distal mean size will become 4.7. If we find a systematic bias in averaging, would it mean that observers rescale each individual size prior to averaging? Not necessarily. It is possible that the observers calculate the mean retinal size and the mean distance (Wardle, Bex, Cass, & Alais, 2012) independently and then multiply one by another.

In Experiment 2, we used a different approach to avoid this problem. Our approach is based on a previously documented property of ensemble statistics that they are sensitive to feature variation among individual items. Specifically, the averaging error tends to grow with the range of feature variation (Fouriezos, Rubenfeld, & Capstick, 2008; Im & Halberda, 2013; Marchant et al., 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014). In other words, the more dissimilar items are present in a set, the less precise observers are at estimating their mean feature. This provides an interesting opportunity to test whether individual items are rescaled, when both the retinal sizes and apparent distances are controlled and only *size-distance correlations* are manipulated. If the retinal sizes of individual objects are positively correlated with their apparent distances, then the apparent size contrast should increase compared to the retinal contrast, because large items at remote planes should appear even larger and small items at close planes should appear even smaller. Negative correlations should reduce an apparent contrast between the small and large items. Therefore, if observers average between individually rescaled items they would show a greater error when adjusting the mean size of a set with positively correlated angular size and apparent distance, and vice versa.

Method

Participants

Twenty-six undergraduate students of the Higher School of Economics (15 female, average age is 20 years) participated in the experiment for extra credits in a psychology course. 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 prior to the experiment. None of the observers that participated in this experiment participated in Experiment 1.

Apparatus, stimuli, and procedure

The apparatus and stimuli used were identical to those used in Experiment 1. The only exception concerned how binocular disparity was manipulated in stereoscopic pairs. Disparity varied between items so that these items could be assigned to several planes. The rule of assignment differed between conditions (Figure 3). In one condition, a *positive correlation* between the angular size and the distance was set via changes in binocular disparity. For two smallest circles, disparity was about –.42º, making them to appear closer than the plane of fixation. For next two circles by size rank, disparity was –.21º. The third ranked pair of sizes had disparity of +.21º. Finally, the fourth ranked pair (two largest circles) had a disparity of +.42º. In another condition, a *negative correlation* between the angular size and the distance was set. Here, two smallest circles were assigned to the +.42º disparity, etc. It is important to note, that this method of size-distance manipulation ensured that (1) angular sizes stayed the same across the positive and negative correlation conditions, (2) apparent distances were also the same, and (3) an expected averaging error within each plane also stayed the same, because each plane always contained two most similar sizes. To obtain the baseline of averaging based solely on the retinal size, a *control* condition was added where all items were assigned a zero disparity.

As the logic of our experiment was based on the assumption that the precision of averaging depends on the range of feature variation, we also manipulated the range of retinal sizes to directly test whether this assumption works for our stimulation. In narrow range trials, individual sizes

varied between 1º and 2.1º. In broad range trials, individual sizes varied between 0.5º and 2.7º degrees. The mean angular size of a set varied 1.1º to 1.8º degrees for both ranges.

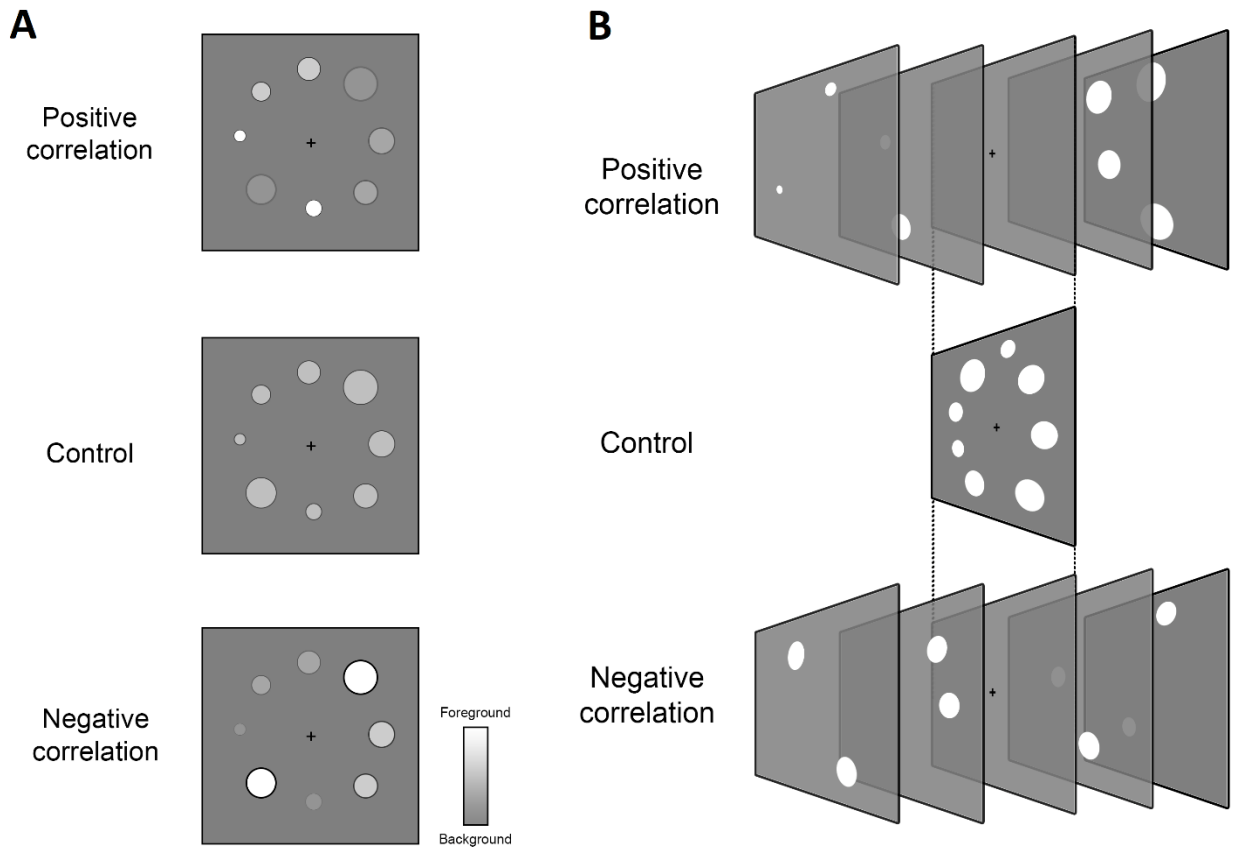

Figure 3. Stimulus views as a function of size-distance correlation in Experiment 2. (A) Fused frontal views of displays having exactly same angular sizes but different distance-size correlations. For illustration purposes, different contrast levels are used to show distance variations and black outlines are used to make low-contrast images more visible to the reader; in real stimulation, all circles had same contrast levels and were uniformly white. (B) The distribution of circles over apparent distances with rescaled sizes.

The procedure of Experiment 2 was the same as the procedure of Experiment 1.

Design and data analysis

Experiment 2 had a 3 (Size-distance correlation: positive vs. negative vs. control) × 2 (Range: narrow vs. broad) within-subject design. For each cell of the design, 50 trials were presented, yielding 300 trials per observer in total. The magnitude of *Error* (see Methods section of Experiment 1, for description) was the principal dependent variable. Two-way repeated-measures ANOVA in both frequentist and Bayesian implementations (Bayesian priors were set the same way as in Experiment 1) was applied to the data.

Results and discussion

We found evidence of an effect of the Size-Distance correlation on the precision of averaging ($F(2,50) = 6.40$, $p = .003$, $\eta^2 = .204$, $BF_{10} = 3.28$). Displays with the negative size-distance correlation yielded a smaller error than displays with the positive correlations ($t(51) = 3.88$, $p < .001$, Bonferroni-Holm corrected $\alpha = .017$, $d = .538$, $BF_{10} = 82.64$) and control displays ($t(51) = 2.41$, $p = .019$, Bonferroni-Holm corrected $\alpha = .025$, $d = .335$, $BF_{10} = 2.11$). We also found a strong effect of the Range ($F(1,25) = 43.33$, $p < .001$, $\eta^2 = .645$, $BF_{10} = 1.59\times10^{51}$) that is provided by a greater error in broad-range displays compared to narrow-range displays (Figure 3A). There was no evidence for an interaction between these factors ($F(2,50) = .786$, $p = .461$, $\eta^2 = .030$, $BF_{\text{main effects only}} / BF_{\text{main effects + interaction}} = 192.5$).

Two important findings were made in Experiment 2. First, we replicated the previously reported detrimental effect of the feature range on the precision of averaging (Fouriezos et al., 2008; Im & Halberda, 2013; Marchant et al., 2013; Maule & Franklin, 2015; Utochkin & Tiurina, 2014), thus directly showing that size variation worked as a limiting factor of precision in our task. Most importantly for our main research question, we found a similar effect elicited by Size-Distance correlation. Specifically, the negative correlation, which is supposed to make an apparent range narrower, yielded a smaller error. This effect is not as dramatic as the effect of the Range (Figure 4). As in Experiment 1, this might be in part due to the range of apparent distances that was not very broad and, hence, provided only a moderate contrast between angular (proximal) and apparent (distal) sizes. We also found no evidence for a symmetrically increasing error in displays with positive Size-Distance correlations compared to control displays. One possible explanation for such an asymmetry can be that the error in size averaging is a decelerating function of the range (Im & Halberda, 2013). The decelerating character of the error, together with the relatively small distance range, could provide a larger contrast between the negative correlation displays and the control displays than between the positive correlation displays and the control displays.

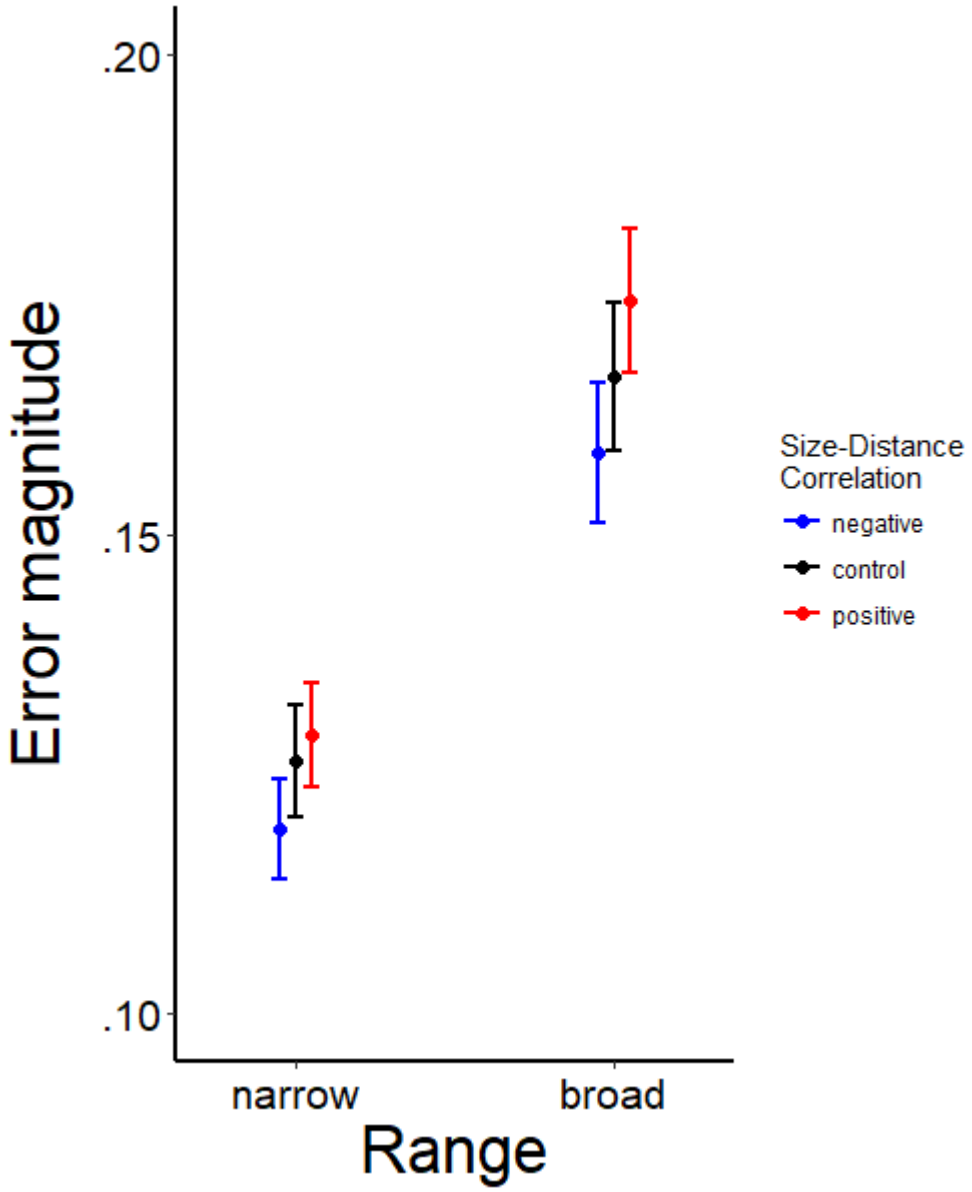

Figure 4. Effects of Range and Size-Distance correlation on the magnitude of the *Error* in Experiment 2. Error bars denote 95% CIs.

Noteworthy, averaging was more precise under negative correlation than under the control condition, despite the items have been located in several different planes in the former case and in a single plane in the latter case. It turns out, therefore, that mean size perception is not impaired when the objects are presented at various distances. Additionally, this finding makes it implausible that the visual system relies on mean distance to infer mean size. As the precision of mean distance (depth) estimation decreases with the physical depth variability (Wardle et al., 2012), we would predict that it should entail the precision of size averaging to fall down as well both in the positive and negative correlation displays. This prediction is not consistent with our pattern. We also can

rule out that observers might average within local subsets grouped by a common plane and then compute the grand mean from these local means. Such a strategy is implausible because it is problematic to compute several means of different groups at one time, as every separate mean representation appears to require the limited capacity of attention and working memory, but it is not a big problem to compute the mean of a single group with any number of elements (Attarha & Moore, 2015a, b; Attarha, Moore & Vecera, 2014). If our observers calculated local means in different planes, then their performance under the negative Size-Distance correlation would have been less precise than under the control condition where all items can be grouped by a single plane. Taken together, our findings that averaging precision is both improved by the reduced range of apparent sizes and not impaired by the increased distance range let us conclude that individual sizes are likely to be rescaled to distances prior to averaging.

General discussion

The results of our study have important implications for both ensemble perception and size constancy. For the field of ensemble perception, our study demonstrated that the summary statistics of multiple sizes can be based on quite elaborated object representations rather than on “raw” retinal measurements of basic features. These representations take into account individual contexts which these features belong to. Specifically, we showed that individual retinal sizes are rescaled to their particular distances, so that an averaged representation is based on these multiple rescaled representations. In addition to Im and Chong's (2009) earlier demonstration of mean size rescaling under the Ebbinghaus illusion, our data show that mean size rescaling clearly supports size constancy in ensemble perception. Another previous study (Yamanashi Leib, Fischer, Liu, Qui, Robertson, et al., 2014) reported evidence for an ability to average between high-level facial features shown from different viewpoints, thus suggesting another example of invariance in ensemble perception. Our findings suggest that ensemble invariance can emerge in lower-level domains of visual processing.

From the viewpoint of size constancy, our results demonstrate that multiple objects can be perceived correctly rescaled at one time. This is a non-trivial conclusion, given that the perception of size invariance is a matter of conditional binding between retinal sizes and distance cues. The existence of the binding problem is especially pronounced when observers are briefly exposed to multiple objects with different features (Cave & Wolfe, 1999; Treisman, 1996), which is the case in our experiments. Indeed, in order to know the rescaled sizes of multiple objects the visual system should compute not only individual retinal sizes and individual distances, but also which distance goes with which retinal size. For combinatorial reasons, the latter task can be very computationally demanding (Tsotsos, 1988) and much harder than the rescaling of one or two objects, as measured in standard constancy experiments. However, we figured out in the present study that conditional binding of this sort is a task that the visual system can do at least satisfactorily. We can add an interesting observation that rescaling seems to be quite irresistible indicating the automatic character of conditional binding (Di Lollo, 2012; Rosenholtz et al., 2012). This conclusion about relative automaticity is based on the principal result of Experiment 2. As the apparent range of sizes that we manipulated via size-distance correlation is irrelevant for size averaging, the range-associated changes in precision suggest that observers likely perform mandatory rescaling.

Our conclusion that size-distance binding and rescaling are performed for multiple objects at one time automatically seems to be in a contradiction with the previously demonstrated role of attention in rescaling (Fang et al., 2008). As Fang and colleagues (2008) showed, diverted attention is associated with reduced activation in brain areas (LOC and PPA) involved in representing high-order pictorial distance and depth cues of the scene, such as linear perspective. Based on this finding, we can infer that the attenuated (though not eliminated) rescaling effect can be explained by the lacking representation of distance information rather than by binding itself. In our experiments, we used binocular rather than pictorial depth cues. The binocular cues based on disparity are supposed to be processed at lower levels of the visual cortex, such as V1 (see review: Blake & Wilson, 2011), where attentional modulation appears to be limited (Luck, Chelazzi,

Hillyard, & Desimone, 1997) or at least to occur later than early feedforward processing takes place (Martínez et al., 1999, 2001; but see Kelly, Gomez-Ramirez, & Foxe, 2008; see also Slotnick, 2017, for review of existing controversies). Given that rescaled sizes also can be represented at V1 (Murray et al., 2006; Ni et al., 2014; Sperandio et al., 2012), binding of binocular distances to retinal sizes can be performed quite early (at least once stereopsis is achieved), as it would not require re-entrant feedback signals from higher-level areas, as in the case of Fang et al. (2008) task.

Another more general consideration of the role of attention in rescaling and averaging follows from the comparison between our work and that by Fang et al. (2008). Fang et al. (2008) observed reduced cortical rescaling under attentional load by a central task complicating attentional spread over the scene (Cohen, Alvarez, & Nakayama, 2011; Joseph, Chun, & Nakayama, 1997). We did not load our observers with any additional task. So, even despite suggesting quite an irresistible character of rescaling multiple objects prior to averaging, our results do not rule out the possible role of attention. Previous research in ensemble perception shows that the capacity to sample ensemble summaries may exceed the known limits of focused attention (Alvarez & Oliva, 2008; Ariely, 2001; Attarha & Moore, 2014; Attarha et al., 2014; Utochkin & Tiurina, 2014) but it is not free of attentional demands (Huang, 2015; Jackson-Nielsen, Cohen, & Pitts, 2017; McNair, Goodbourn, Shone, & Harris, 2017). It is probably more correct to speak of globally distributed rather than narrowly focused attention here (Alvarez, 2011; Treisman, 2006). It is possible that the rescaling of multiple sizes and subsequent averaging are also carried out under distributed attention.

Having demonstrated efficient conditional binding for multiple sizes and distances, we are cautious about claims that this finding can be automatically generalized for any other couplings of features. Explicit doubts were expressed by Treisman and colleagues (Chong & Treisman, 2005; Emmanouil & Treisman, 2008; Treisman, 2006) that correct item-by-item feature binding is a general case in ensemble perception. At the same time, there is evidence that some feature

conjunctions can be processed by selectively tuned units similar to those tuned to separate features (Holcombe & Cavanagh, 2001; Livingstone & Hubel, 1988; Seymour, Clifford, Logothetis, & Bartels, 2010; Zhaoping & Zhe, 2012), thus removing the need for the binding "bottleneck" (Di Lollo, 2012) and making binding automatic (Rosenholtz et al., 2012). We suppose that size-distance binding observed in our study can be also based on a neural mechanism of this sort.

Conclusion

To summarize, our study suggests that ensemble summary statistics of size works as more than a coarse "sketch" of the visual scene. We figured out that, when computing ensemble summaries, the visual system takes into account nuances of the context – individual distances – which ensemble constituents appear in. It appears that an ability to maintain the satisfactory level of perceptual constancy for numerous objects at once supports this property of ensemble statistics. As a result, ensemble representation seems to be a perceptual approximation of the physical environment rather than of the retinal image. As such, our study adds to our understanding of how visual summary statistics support the perceived richness of the world under severe limitations of deep object processing (Cohen, Dennett, & Kanwisher, 2016). Ensemble mechanisms provide sufficiently elaborated information for numerous objects, that we cannot focus on, to look closer to their physically realistic features.

Acknowledgements

The reported study was funded by RFBR according to the research project #18-313-00253. The authors thank Chris Oriet and two anonymous reviewers for their helpful comments on earlier drafts of the manuscript. The authors also thank Yuri A. Markov for help with the design of figures and support throughout writing this manuscript.

Open science statement

Experimental scripts and data can be publicly accessed via Open Science Framework at https://osf.io/t9e6q/

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. http://doi.org/10.1016/j.visres.2013.02.018

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131. http://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. http://doi.org/10.1111/j.1467-9280.2008.02098.x

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science : 12*(2), 157–162. http://doi.org/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 & Psychophysics, 70*(7), 1325–1326. http://doi.org/10.3758/PP.70.7.1325

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

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1440–1449. http://doi.org/10.1037/a0036206

Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record, 59*(2), 171–186.

Bauer, B. (2017). Perceptual averaging of line length: Effects of concurrent digit memory load. *Attention, Perception, & Psychophysics, 79*(8), 2510–2522. http://doi.org/10.3758/s13414-017-1388-4

Blake, R., & Wilson, H. (2011). Binocular vision. *Vision Research, 51*(7), 754–770. http://doi.org/10.1016/j.visres.2010.10.009

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384–392. http://doi.org/10.1177/0956797610397956

Brand, J., Oriet, C., & Sykes Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology, 66*(1), 63-69. doi: 10.1037/a0026950

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

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision, 17*(2), 21. http://doi.org/10.1167/17.2.21

Chong, S. C., Joo, S. J., Emmmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70*(7), 1327–1334. http://doi.org/10.3758/PP.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404. http://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. http://doi.org/10.1016/j.visres.2004.10.004

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

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science, 22*(9), 1165–1172. http://doi.org/10.1177/0956797611419168

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

Corbett, J. E. (2016). Gestalt grouping facilitates perceptual averaging to increase the efficiency of memory representations. *Journal of Vision, 16*(12), 305-305.

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. http://doi.org/10.3758/s13414-013-0595-x

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24*(1), 87–185. http://doi.org/10.1017/S0140525X01003922

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. http://doi.org/10.1016/j.tics.2012.04.007

Duwaer, A. L., & Van Den Brink, G. (1981). What is the diplopia threshold? *Perception & Psychophysics, 29*(4), 295–309. http://doi.org/10.3758/BF03207338

Ehrenstein, W. H., & Hamada, J. (1995). Structural factors of size contrast in the Ebbinghaus illusion. *Japanese Psychological Research, 37*, 158–169. https://doi.org/10.4992/psycholres1954.37.158

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

Epstein, M. L., & Emmanouil, T. A. (2017). Ensemble coding remains accurate under object and spatial visual working memory load. *Attention, Perception, & Psychophysics, 79*(7), 2088–2097. http://doi.org/10.3758/s13414-017-1353-2

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. http://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. http://doi.org/10.3758/PP.70.3.456

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision, 14*(9), 22. http://doi.org/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. http://doi.org/10.1016/j.cub.2007.06.039

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience, 4*(2), 127–128. http://doi.org/10.1038/83945

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

Huang, L. (2015). Statistical properties demand as much attention as object features. *PLOS ONE,* 10(8), e0131191. http://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*(2), 375 - 384. https://doi.org/10.3758/APP.71.2.375

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. http://doi.org/10.3758/s13414-012-0399-4

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

Jaeger, T. (1978). Ebbinghaus illusions: Size contrast or contour interaction phenomena? *Perception & Psychophysics*, *24*(4), 337–342. https://doi.org/10.3758/BF03204250

Jaeger, T., Klahs, K., & Newton, D. (2014). Ebbinghaus illusions with disc figures: Effects of contextual size, separation, and lightness. *Perceptual and Motor Skills*, *118*, 805-817. http://doi.org/10.2466/24.27.PMS.118k29w0

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

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature*, *387*(6635), 805–807. http://doi.org/10.1038/42940

Julesz, B. (1971). *Foundations of cyclopean perception*. University of Chicago Press.

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. http://doi.org/10.1163/156856806778457377

Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cerebral Cortex*, *18*(11), 2629–2636. http://doi.org/10.1093/cercor/bhn022

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

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

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, *240*(4853), 740–749. http://doi.org/10.1126/science.3283936

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. http://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), 279–81. http://doi.org/10.1038/36846

Mack, A., & Rock, I. (1998). *MIT Press/Bradford Books series in cognitive psychology. Inattentional blindness.* Cambridge, MA: The MIT Press. http://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. http://doi.org/10.1016/j.actpsy.2012.11.002

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

Martínez, A., Di Russo, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: Timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research, 41*(10–11), 1437–1457. http://doi.org/10.1016/S0042-6989(00)00267-4

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

Mckee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research, 32*(8), 1447–1460. http://doi.org/10.1016/0042-6989(92)90201-S

McNair, N. A., Goodbourn, P. T., Shone, L. T., & Harris, I. M. (2017). Summary statistics in the attentional blink. *Attention, Perception, & Psychophysics, 79*(1), 100–116. http://doi.org/10.3758/s13414-016-1216-2

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. http://doi.org/10.1037/h0043158

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. http://doi.org/10.1038/nn1641

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

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. http://doi.org/10.1016/j.cub.2014.06.003

Oriet C. & Brand J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research, 79*, 8–16. 10.1016/j.visres.2012.12.004

Palmer, D. A. (1961). Measurement of the horizontal extent of panum's area by a method of constant stimuli. *Optica Acta: International Journal of Optics, 8*(2), 151–159. http://doi.org/10.1080/713826374

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

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

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

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception, 34*, 847–856. doi: 10.1068/p5273

Robinson, J. O. (1998). *The Psychology of Visual Illusion.* Dover Publications, Mineola, New York.

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

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*. http://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. http://doi.org/10.3758/PBR.16.2.225

Seymour, K., Clifford, C. W. G., Logothetis, N. K., & Bartels, A. (2010). Coding and binding of color and form in visual cortex. *Cerebral Cortex, 20*(8), 1946–1954. http://doi.org/10.1093/cercor/bhp265

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

Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics, 70*(7), 1335–1336. http://doi.org/10.3758/PP.70.7.1335

Slotnick, S. D. (2017). The experimental parameters that affect attentional modulation of the ERP C1 component. *Cognitive Neuroscience*, 1–10. http://doi.org/10.1080/17588928.2017.1369021

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. http://doi.org/10.1038/nn.3069

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

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

Treisman, A., & Gelade, G. (1980). A Feature-Integration of Attention. *Cognitive Psychology, 136*, 97–136.

Tsotsos, J. (1988). A “complexity level” analysis of immediate vision. *International Journal of Computer Vision, 1*(4), 303-320. http://doi.org/10.1007/BF00133569

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. http://doi.org/10.1016/j.actpsy.2013.11.012

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

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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F.,Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1–19. http://doi.org/10.3758/s13423-017-1323-7

Wardle, S. G., Bex, P. J., Cass, J., & Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *Journal of Vision, 12*(6), 12–12. http://doi.org/10.1167/12.6.12

Watamaniuk, S. N., & 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. http://doi.org/10.3758/s13423-011-0071-3

Whitney, D. & Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69,* 10.1146/annurev-psych-010416-044232.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24*(1), 11–17. http://doi.org/10.1016/S0896-6273(00)80818-1

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

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

Zhaoping, L., & Zhe, L. (2012). Properties of V1 neurons tuned to conjunctions of visual features: application of the V1 saliency hypothesis to visual search behavior. *PLOS ONE, 7*(6), e36223. http://doi.org/10.1371/journal.pone.0036223