

Ensemble size judgments account for size constancy

Jason Haberman & Sneha Suresh

**Attention, Perception, &
Psychophysics**

ISSN 1943-3921
Volume 83
Number 3

Atten Percept Psychophys (2021)
83:925-933
DOI 10.3758/s13414-020-02144-6

Your article is protected by copyright and all rights are held exclusively by The Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Ensemble size judgments account for size constancy

Jason Haberman¹ · Sneha Suresh²Accepted: 9 September 2020 / Published online: 20 October 2020
© The Psychonomic Society, Inc. 2020

Abstract

The natural environment is full of redundant information that the visual system compresses into an ensemble representation by averaging features of groups of items. Ensemble perception has been shown to operate with remarkable flexibility, efficiently integrating information across a variety of visual domains. In the current set of experiments, we tested whether average size representations reflect the physical size of objects displayed on a screen or perceptual transformations due to size constancy. We induced a perceptual change by presenting sets of triangles with linear perspective cues – lines converging at the horizon. Assuming a constant size, these cues cause individual objects “in the distance” to appear larger than objects without distance cues, due to size constancy heuristics. Observers viewed sets of triangles with and without linear perspective cues and judged whether a subsequently presented test triangle was larger or smaller than the average size of the preceding set. Results revealed ensemble size representations took size constancy into account, reflecting the perceived size of the triangles rather than their absolute size. Interestingly, the amount of bias exhibited was well characterized by the summed bias associated with each of the three triangles presented individually. Other pictorial cues to depth, such as occlusion and height-in-field, did not elicit the same bias when those were the only depth cues available. Overall, our results complement and extend other work showing that average size reflects the perceptual size of individual items in a set.

Keywords 3D perception: depth and shape from X · Object recognition · Perceptual organization · Ensemble Perception

Introduction

The visual system possesses an extensive array of heuristics to assist in the reconstruction of the natural environment. Among the most impressive and contentious of these is the phenomenon of size constancy (Carlson, 1962; Ross & Plug, 1998), which refers to the consistent perception of an object's size despite changes in perceived distance and retinal image size. A tiny retinal image could reasonably be perceived as a tiny

object, but the visual system incorporates perceived distance information to derive the correct interpretation of size. Holway and Boring (1941) originally introduced the size-distance invariance hypothesis, but whether the rescaling of object size is veridical depends on the perspective an observer is asked to take when viewing a scene and the nature of the task (e.g., Epstein, Park, & Casey, 1961; Kilpatrick & Ittelson, 1953). Regardless of the precision of size rescaling, it is unambiguous that the visual system rescales an object's perceived size based on distance information.

Size constancy intimately depends on the perceived distance of an object, which may be inferred from several different cues. Binocular cues, such as disparity and Descartes' convergence, depend on the unique positioning of each eye in the head. Because there are cells in V1 explicitly sensitive to disparity information (Barlow, Blakemore, & Pettigrew, 1967), it is thought that stereopsis arises early within the visual hierarchy. Distinct pictorial or monocular cues to perceived distance, including linear perspective and height-in-field (and many others), provide compelling depth information even when depicted on a two-dimensional surface (Berkeley, 1709). These cues are thought to emerge at various points throughout the visual hierarchy (Landy, Maloney, Johnston, & Young, 1995) and are not relegated to early processing in

Significance Ensemble perception is a ubiquitous visual phenomenon. While much research has focused on defining its limitations, this work examines its nature. For the domain of ensemble size, we ask whether it reflects absolute, retinal image size or perceived size after rescaling has occurred? Our work reinforces the notion that ensemble size takes into account size transformations due to depth estimates and is the first demonstration of its kind using pictorial cues to depth.

✉ Jason Haberman
habermanj@rhodes.edu

¹ Department of Psychology, Rhodes College, 2000 North Parkway, Memphis, TN 38112, USA

² Brigham and Women's Hospital, Harvard University, Cambridge, MA, USA

the way that binocular information is. Thus, the visual system has developed multiple, independent routes to extracting distance information from a scene, which are used to create size constancy.

Ensemble perception is a grouping heuristic that compresses redundant visual information into a representative statistic, such as the mean or variance (e.g., Ariely, 2001; Dakin & Watt, 1997; Solomon, 2010; Watamaniuk & Duchon, 1992). It reconciles the grand illusion (Noe, Pessoa, & Thompson, 2000), accounting for the strong impression of visual completeness despite the numerous demonstrations of a consciousness bottleneck (Luck & Vogel, 1997; Rensink, O'Regan, & Clark, 1997; Simons & Chabris, 1999). Ensembles are efficient (Brady, Shafer-Skelton, & Alvarez, 2017; Chetverikov, Campana, & Kristjánsson, 2017; Haberman & Whitney, 2011; Jazayeri & Shadlen, 2010; Olkkonen, McCarthy, & Allred, 2014) and robust to interference (Chong & Treisman, 2005; Haberman & Ulrich, 2019; Sweeny & Whitney, 2014), come online prior to individual object recognition (Sweeny, Wurnitsch, Gopnik, & Whitney, 2015), and operate independently at multiple levels within the visual hierarchy (Haberman, Brady, & Alvarez, 2015). Given their ubiquity, it is critical to probe the nature of ensemble representations. Do they represent a straightforward, linear combination of the physical inputs, or are ensembles conceptually transformed to represent the ultimate conscious percept? While many early computations within the visual hierarchy do not reflect visual consciousness (e.g., He, Cavanagh, & Intriligator, 1996), some evidence suggests that the computations associated with object size rescaling are represented as early as V1 (Murray, Boyaci, & Kersten, 2006); retinotopic activation is more widespread than would be expected based on the incoming angular size of a stimulus, although these activations manifest as a result of feedback mechanisms. Although the visual system represents many ensemble domains, given object size perception is subject to rescaling, average size perception provides an appealing area for exploration.

Some researchers have examined the issue of ensemble-size rescaling (Im & Chong, 2009; Tiurina & Utochkin, 2019). Im and Chong (2009) presented observers with sets of target circles flanked by large or small distracters, which induced the well-known Ebbinghaus Illusion. Under typical viewing conditions, flanking circles alter the perceived size of a central target circle (larger flankers make a target seem smaller than another identical target surrounded by small flankers). In their experiments, researchers presented multiple “Ebbinghaus sets,” where observers had to compare the average size of the target circles on one side of the screen with those on the other side. Thus, they were testing whether ensemble size perception reflected the physical size of the target circles or their perceived size, rescaled in the presence of the flankers. Their results confirmed the latter; observers were unlikely to report the average size of a set of targets surrounded

by large flankers as bigger than a set of targets surrounded by small flankers, even though the average sizes were identical.

More recently, Tiurina and Utochkin (2019) used a clever design to further assess whether ensemble size perception accounts for size constancy. Researchers presented sets of circles of varying sizes stereoscopically, such that sets appeared in front of, behind, or at the plane of fixation. The reported average size of circles that stereoscopically appeared beyond the plane of fixation was larger than the average size at or in front of the plane of fixation, consistent with size rescaling for individual objects. In a follow-up experiment, researchers asked observers to report the average size of a set of circles that appeared on multiple depth planes. On half the trials, there was an inverse relationship between the retinal image size and perceived depth, meaning objects seen as further away had a smaller angular size. This relationship is consistent with Emmert's Law, and serves to effectively reduce the perceived range of sizes within the set, which increases ensemble representation precision (Im & Halberda, 2013; Maule & Franklin, 2015). On the other half of trials, the relationship was reversed – objects perceived as farther away cast a *larger* retinal image. Error in average-size judgments was smaller when perceived distance and retinal image size were inversely related. Given that result, researchers concluded individual object size must be rescaled prior to averaging.

As mentioned, perceived depth may be influenced by many factors, which then affect perceived size. In their experiments, Tiurina and Utochkin (2019) manipulated binocular cues to depth, which arise early in the visual-processing stream (Barlow et al., 1967). In the current set of experiments, we manipulated pictorial cues to depth, which arise at various points along the visual hierarchy. In Experiment 1, for example, we created an impression of depth by presenting sets of triangles in the context of linear perspective cues. Extracting depth from linear perspective cues must be learned (Segall, Campbell, & Herskovits, 1966), and thus might require more exposure to perceive than binocular disparity. Thus, we explored whether average size perception reflects the perceived size of objects rescaled in response to monocular, pictorial cues to depth. To preview our results, we found complementary results to those of previous researchers (Im & Chong, 2009; Tiurina & Utochkin, 2019), showing that the average size of distally perceived triangles is greater than the average of triangles viewed without any cues to depth (Experiment 1). Interestingly, the extent to which *individual* items are rescaled closely matches the extent to which the perceived ensemble is rescaled (Experiment 2), confirming the notion that ensemble representation follows size constancy, and is, in fact, highly dependent on the individual items. Finally, not all pictorial cues to depth are equally effective – impoverished scenes showing only either occluded stimuli or height-in-field stimuli did not reveal ensemble-size rescaling (Experiment 3).

Experiment 1

Methods

Participants

Sixteen Rhodes College undergraduates, aged 18–22 years, participated in this study for either course credit or monetary compensation.

The compensation rate was course credit or \$10/h. All participants gave informed consent and had normal or corrected-to-normal vision. This research, and all research described herein, was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board at Rhodes College.

Stimuli and design

Observers were seated 63 cm away from a 19-in. Dell flat screen monitor (60-Hz refresh rate) in a dimly lit room, with their head placed on a chin rest. Experiments were implemented using custom scripts written in Psychtoolbox version 3.0.11, (Brainard, 1997) within MATLAB (R2015a The MathWorks Inc., Natick, MA, USA). Observers viewed sets of three isosceles triangles arranged in one of two conditions: either with or without linear perspective cues (Fig. 1). The linear perspective cue consisted of two lines, 47 cm and 32 cm in length, appearing to converge in the distance in a northwesterly direction. The triangles were arranged obliquely within the two lines (in a manner similar to the *Corridor Illusion*). In the second condition, which had no linear perspective cues, the triangles were arranged along the horizontal meridian without any lines. This arrangement was chosen to reduce any impression of depth provided by the obliquely arranged

triangles, which might occur even without the physical lines. Following the set, observers were presented with a single test triangle varying in size, presented in the middle of the screen.

On each trial, each of the three triangles within a set had a randomly selected area subtending between 3.9° and 8.8° in triangle width, and 3.4° and 7.6° in triangle height, with a constant aspect ratio across triangles. Once the individual triangle sizes were selected, the size of the subsequent test triangle ranged from $\pm 88\%$ of the average area of the preceding set (i.e., method-of-constant stimuli with incremental steps of 22%, for a total of ten possible relative sizes – note that there were twice as many trials where the test triangle was equal in area to the average size than any other size condition). Although other measures of size, such as average diameter, are viable metrics (e.g., Raidvee, Toom, Averin, & Allik, 2020), the arithmetic area is a more straightforward assessment of average size due to the fact that our stimuli are not equilateral triangles. Additionally, Chong and Treisman (2003) revealed that the representation of the average size of circles likely falls somewhere between the average area and average diameter, making area a reasonable metric for average size.

Procedure

For each trial, observers viewed the triangles with or without the linear perspective cue for 1,000 ms, randomly, followed by a test triangle of varying size for 1,000 ms, and finally the response screen until observer response (Fig. 1). Observers were asked to report whether the test triangle was smaller or larger than the average size of the preceding set (observers indicated their responses using a keypress: ‘F’ for smaller and ‘J’ for larger).

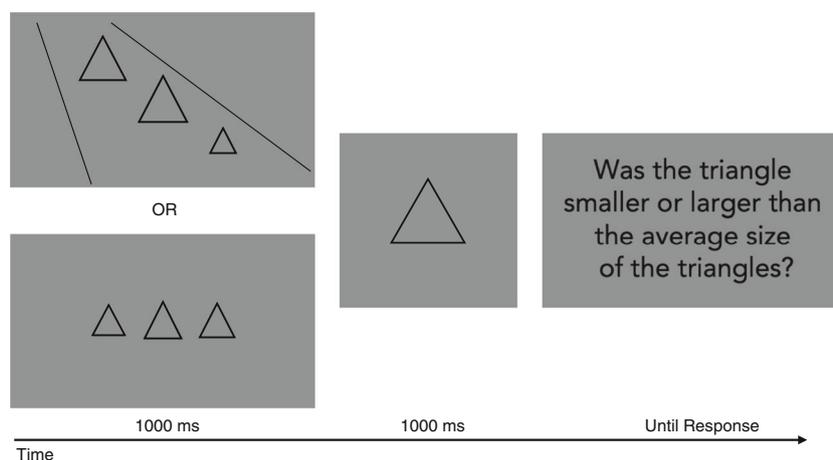


Fig. 1 Trial sequence for Experiment 1. Observers had to indicate whether the test triangle was larger or smaller than the average size of the preceding set

There were 18 trials in each condition (linear perspective vs. no linear perspective, ten relative sizes), for a total of 360 trials. Prior to beginning of the experiment, observers performed 20 practice trials with both conditions, which were discarded from further analysis.

Results and discussion

We calculated the proportion of the time observers reported the test triangle as larger or smaller than the average size of the preceding set. Using the fit function implemented in MATLAB, each observer's point of subjective equality (PSE) was determined from a logistic psychometric function. Any observer whose PSE was 2.5 standard deviations from the average was excluded from analysis. This resulted in the exclusion of two observers, leaving a total of 14 participants.

Figure 2 depicts the data for one representative observer. The x-axis depicts the percent size difference between the test triangle and the average size of the preceding set, and the y-axis depicts the proportion of trials on which this observer reported the test triangle as larger than the average of the preceding set. When the linear perspective cues were visible, observers were less likely to report the test triangle as larger, indicated by the rightward shift of the psychometric function. This suggests that observers viewed the average size of the sets with linear perspective cues as larger than the average size of the sets without such cues. This trend emerged at the group level, as determined by a within-subjects t-test comparing PSEs between conditions ($M_{\text{perspective}} = 0.06$; $M_{\text{no perspective}} = -0.09$; $t(13) = 7.91$, $p < 0.001$, $d = 2.11$). These results suggest observers take perceived size into account when estimating the average size of a set of items, placing the ensemble calculus beyond that of size constancy.

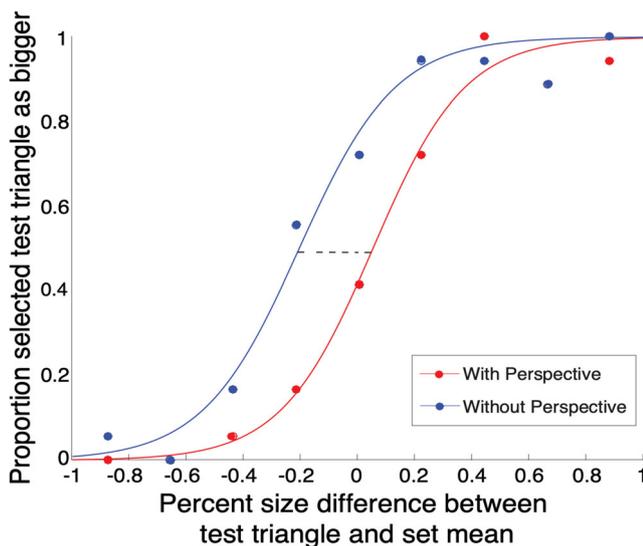


Fig. 2 Psychometric functions for one representative observer. When linear perspective cues were present, observers were less likely to report the test triangle as larger than the average

Experiment 2

In judging the effect of the linear perspective cues on average size judgments, the previous experiment did not account for how each individual item might differentially contribute to the bias. Experiment 2 explicitly measured the extent to which each individual item was susceptible to a shift due to size constancy adjustments.

Methods

Participants

Thirty-nine Rhodes College undergraduates, aged 18–22 years, participated in this study for either course credit or monetary compensation.

Stimuli and design

Nearly all stimulus and design elements were as described in Experiment 1. However, instead of judging the average size of a set of three triangles, observers judged the size of just a single, cued triangle, with or without a linear perspective cue, in one of three positions (Fig. 3). The triangle position could be top, middle, bottom for linear perspective or left, middle, right for no linear perspective. All three triangles were present on every trial (as in Experiment 1), but observers were cued to just one of the triangles by a green frame. Observers were randomly assigned to one of the three positions for the duration of an experimental session, which left 13 observers for each condition.

Procedure

The task was nearly identical to the task described in Experiment 1. Instead of judging the average size of three triangles, observers judged the size of just a single triangle (Fig. 3). There were 18 trials in each condition (2 perspectives \times 10 relative sizes), for a total of 360 trials for each observer. Prior to starting the experiment, observers performed 20 practice trials both with and without linear perspective cues, which were discarded from further analysis.

Results and discussion

Analysis was similar to that described in Experiment 1, which used PSEs as the dependent measure. Figure 4 depicts the average PSE difference between linear perspective and no linear perspective conditions for each of the three triangle positions. Larger numbers on the y-axis indicate a larger effect of linear perspective cues. Although triangle position was not significant ($F(2, 36) = 2.59$, $p = .09$, $\eta^2 = 0.13$), the general pattern and effect size suggest observers were more likely to report the size of a

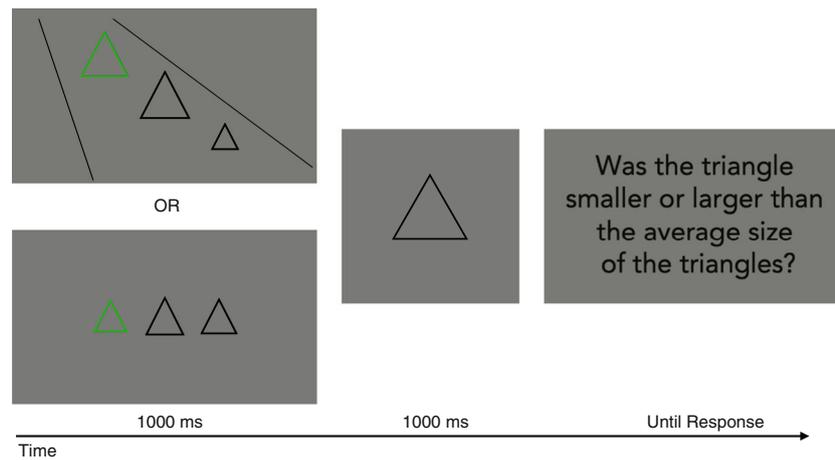


Fig. 3 Trial sequences for Experiment 2. Observers saw one of three triangle positions cued with a green frame and had to indicate whether the test triangle was larger or smaller than that single item

single triangle as larger when it appeared farther away from the observer (Fig. 4). The triangle closest to the observer showed almost no size rescaling. Interestingly, the summed PSE difference of the three triangles individually (0.16) was closely aligned with the PSE difference in ensemble bias observed in Experiment 1 (0.15). This suggests that size rescaling occurs at the individual item level prior to averaging, consistent with the conclusions of Tiurina and Utochkin (2019). This provides further evidence that the representation of the ensemble is deeply dependent on the representation of the individual items (Haberman, Brady, & Alvarez, 2015), but that the efficiency with which information from those items is combined (Whitney & Yamanashi Leib, 2018) allows precise ensemble representations to emerge (Alvarez, 2011).

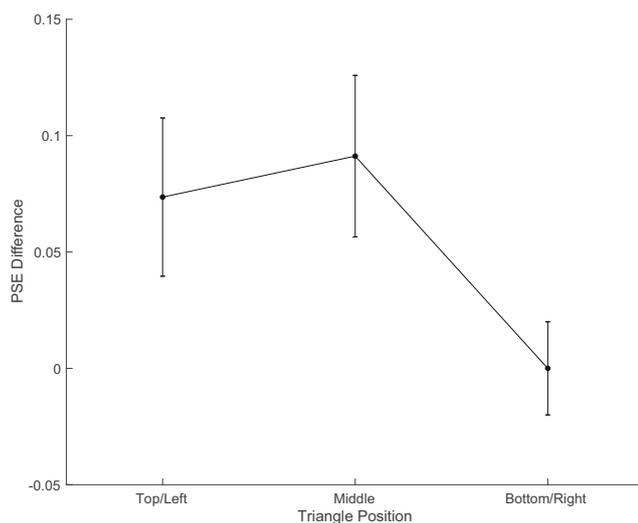


Fig. 4 Average point of subjective equality (PSE) differences for the cued triangle (with vs. without linear perspective cues). Error bars represent one standard error of the mean (SEM)

Experiments 3A and 3B

Although linear perspective is a powerful cue to depth, the visual system uses numerous other cues to extract depth information from a scene. Given the results of Experiment 1, which revealed ensemble size estimates incorporate size constancy, we explored the generalizability of this phenomenon by testing other cues to depth, such as occlusion and height-in-field (Fig. 5).

Occlusion is a powerful cue to depth. Although monocular occlusion does not provide information regarding the magnitude of depth, it unambiguously provides the order in which objects appear (Ono, Shimono, & Shibuta, 1992). Recent ensemble work has revealed robust summary information remains available despite the reduced visibility of stimuli under occluded viewing conditions (Haberman & Ulrich, 2019).

Height-in-field also provides important depth information (Loomis, 2001; Ooi, Wu, & He, 2001). Objects higher in an observer’s field-of-view have a greater perceived distance, appearing to reside closer to the horizon. Here we ask whether either of these depth cues also result in a bias to perceive the average size of triangles seen further away as larger than those seen closer.

Experiment 3A: Occlusion

Methods

Participants

Ten Rhodes College undergraduates, aged 18–22 years, participated in this study for either course credit or monetary compensation. The compensation rate was course credit or \$10/h. All participants gave informed consent and had normal or corrected-to-normal vision.

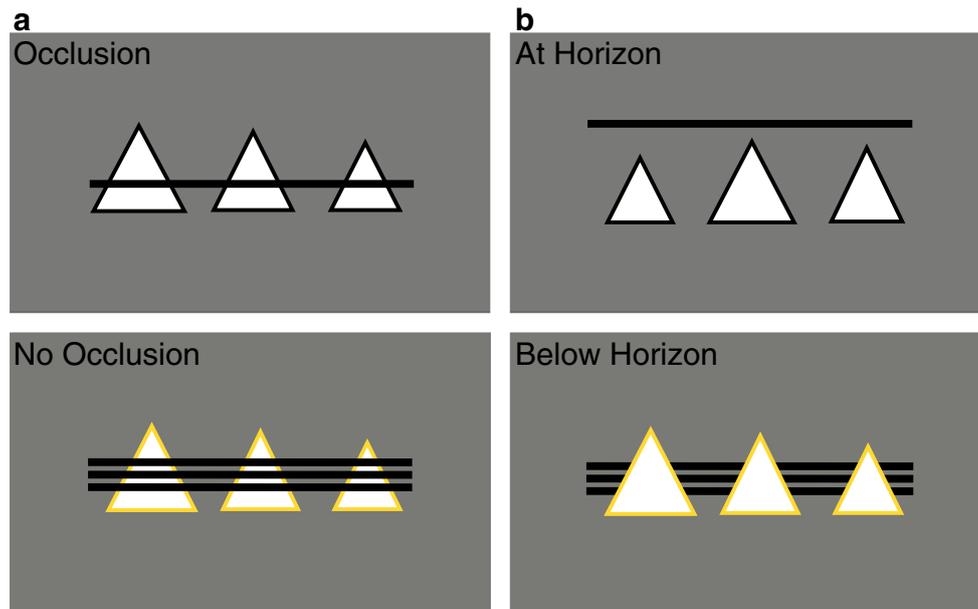


Fig. 5 Example stimuli from Experiment 3. (A) Occlusion; (B) Height-in-field

Stimuli and design

Stimuli and design protocols were identical to Experiment 1, except that observers viewed sets of triangles presented horizontally either behind or in front of a set of three occluding horizontal bars (Fig. 5A). Triangles were surrounded by a yellow frame to increase their visibility. As we were testing specifically the role of occlusion in ensemble-size perception, other depth cues that might typically be present (e.g., linear perspective) were removed.

Procedure

The procedure was identical to Experiment 1, except observers viewed sets of triangles in front of horizontal bars or triangles amodally completing behind those bars (Fig. 5A).

There were 18 trials in each condition (2 perspectives \times 10 relative sizes), for a total of 360 trials. Prior to starting the experiment, observers performed 20 practice trials both with and without linear perspective cues, which were discarded from further analysis.

Experiment 3B: Height-in-field

Participants

Ten Rhodes College undergraduates, aged 18–22 years, participated in this study for either course credit or monetary compensation. The compensation rate was course credit or \$10/h. All participants gave informed consent and had normal or corrected-to-normal vision.

Stimuli and design

Stimuli and design protocols were identical to Experiment 1, except that observers viewed sets of triangles presented horizontally either in line with or below a horizontal line near the center of the screen (Fig. 5A). As we were testing specifically the role of height-in-field in ensemble size perception other depth cues that might typically be present (e.g., linear perspective) were removed.

Procedure

The procedure was identical to Experiment 1, except observers viewed sets of triangles in front of a single horizontal bar or triangles well below the horizontal bar (Fig. 5B).

There were 18 trials in each condition (2 perspectives \times 10 relative sizes), for a total of 360 trials. Prior to starting the experiment, observers performed 20 practice trials both with and without linear perspective cues, which were discarded from further analysis.

Results and discussion

Analysis was done as described in Experiment 1. The PSEs across both conditions failed to reject the null hypothesis, as revealed by two within-subjects t-tests (run separately for Experiments 3A and 3B; $M_{\text{occlusion}} = 0.077$, $M_{\text{no occlusion}} = 0.071$, $t(9) = 0.31$, $p = 0.76$, $d = 0.1$; $M_{\text{horizon}} = 0.129$, $M_{\text{below horizon}} = 0.1$, $t(9) = 1.41$, $p = 0.19$, $d = 0.45$). These results are inconsistent with those of Experiment 1, and suggest these depth cues were less effective at eliciting an impression of depth. While pictorial and binocular cues all provide

some depth information in a scene, they are not all dependent on the same mechanisms, and their individual contributions to depth perception may vary. Occlusion, for example, may arise early within the visual hierarchy, with occlusive T-junctions being a feature of end-stopped cells (Heitger, Rosenthaler, Von Der Heydt, Peterhans, & Kübler, 1992). On its own, occlusion provides only ordinal information. Depth cues, regardless of whether they provide ordinal or metric information, are combined to create more veridical estimates of depth (Burge, Peterson, & Palmer, 2005; Landy et al., 1995). Without additional contextualizing information, impoverished stimuli such as these may be limited in their capacity to elicit an impression of depth, which may limit any rescaling of size that can occur. Additionally, the likelihood of ensemble-size rescaling may be further reduced due to the very nature of the objects used. Observers viewed sets of triangles, which may not have a well-defined prototypical size. Object knowledge, while not critical for the perception of depth, most certainly influences it (e.g., faces are convex, and as such often are perceived in the foreground of a scene; Bertamini, Martinovic, & Wuerger, 2008).

General discussion

These results further support the notion that ensemble size representations reflect perceived size rather than absolute angular size. When observers viewed sets of triangles presented with linear perspective cues, they were biased to perceive the set as overall larger than when those sets were presented along a flat plane without depth cues. This adds to a growing body of literature suggesting ensembles operate later within the visual hierarchy than might be expected given the stimulus domain (Im & Chong, 2009; Tiurina & Utochkin, 2019). In particular, these experiments reveal that ensemble-size rescaling occurs when viewing the monocular, pictorial depth cue of linear perspective, whereas previous work (Tiurina & Utochkin, 2019) showed rescaling in response to binocular depth cues.

We also demonstrated that the extent to which individual items are rescaled predicts the extent to which the ensemble is rescaled. Experiment 2 revealed that objects perceived farther away (top and middle triangles) likely contributed more to the ensemble rescaling than objects seen as closer. This is consistent with a recent ensemble-saliency model proposed by Kanaya, Hayashi, and Whitney (2018), which demonstrated larger (more salient) objects were weighted more heavily in estimates of the mean size than other objects. If objects perceived as farther away are rescaled to a greater extent than closer objects, the saliency model predicts they would contribute more to the perception of mean size, thereby amplifying mean size estimates. Our results support this idea.

Importantly, our work shows that size constancy, induced by linear perspective, is incorporated into ensemble size estimates.

These results challenge the notion that ensemble size perception occurs at a singular locus of processing, as there are now multiple demonstrations showing average size extraction in the presence of several different types of depth cues. Depth perception is a highly interactive process, whereby different depth cues may be extracted modularly, but are ultimately weighted and combined differentially depending on the presence of other cues in the scene (Landy et al., 1995). Our experiments show that, when presented in isolation, a linear perspective cue to depth elicits size rescaling of the ensemble (Experiment 1), while occlusion and height-in-field do not (Experiment 3). However, this does not rule out the possibility that these latter cues may also strongly influence ensemble-size rescaling when presented within the context of a more naturalistic scene, where the information provided by the cues may be properly weighted in accordance with the rest of the cues.

While pictorial cues to depth may differentially contribute to size rescaling, it was nonetheless surprising that height-in-field cues did not elicit a significant bias in Experiment 3B. Precise distance information may be recovered from the angular declination of a target relative to the horizon (Ooi, Wu, & He, 2001). In this sense, height-in-field should provide distance information akin to that provided by linear perspective cues. One explanation for the discrepancy across experiments is that the strength of our particular stimulus implementation was not explicitly equated between the linear perspective cues and the height-in-field cues. The latter is rarely encountered in isolation and requires clear delineation between the ground plane and the horizon (Arterberry, 2008). In an effort to isolate height-in-field from other depth cues, our stimuli were extremely impoverished (Fig. 5B), which may have minimized potential size rescaling effects.

As for occlusion (Experiment 3A), the lack of ensemble-size rescaling may be less surprising. Unlike linear perspective and height-in-field cues, occlusion only provides ordinal depth information. An observer can only recover the order in which objects are displayed and not ratio distance metrics, per se. As such, it is plausible that occluded objects presented on a two-dimensional screen offered little beyond which object was in front of the other.

As mentioned, Experiment 2 suggests the ensemble size calculus may be predicted by the individual item representations, consistent with prior work in other ensemble domains (Alvarez & Oliva, 2008; Haberman, Brady, & Alvarez, 2015). The overall amount of bias seen when observers judged the average size of a whole set was closely aligned with the summed bias for individual item judgments. While this does not discount the numerous studies demonstrating the efficiency of ensemble representations (e.g., Alvarez, 2011; Chong & Treisman, 2003; Haberman & Whitney, 2010, 2011), it does reinforce the notion that the ensemble representation depends on the quality on the inputs (Haberman et al., 2015). Although the specific position of the triangle did not relate to the extent

of rescaling, our pattern of results is consistent with that of Tiurina and Utochkin (2019), which suggests items perceived further away from the plane of fixation are rescaled to a greater degree than those perceived at the plane of fixation, and are done so before the ensemble calculus occurs.

A number of ensemble experiments point to the interesting disconnect between the quality of individual item representations, typically poor, and the quality of the ensemble representations, typically high-fidelity (e.g., Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014; Ariely, 2001; Fischer & Whitney, 2011; Haberman & Whitney, 2011). In our experiments, observers had a full second to examine the sets, which arguably is sufficient time to scrutinize the individual items. It would be interesting to explore in a future study whether ensemble-size rescaling requires this longer viewing time, or whether such effects might still emerge given shorter stimulus exposure.

Overall, our results point to the complex nature of ensemble representations. For average size, the representation appears to be based on the perceived size of individual objects rather than their absolute angular size. This occurred despite the somewhat ambiguous objects used, triangles, which do not possess a prototypical size. Given the varied ways in which depth is perceived (e.g., binocularly, pictorially), and that ensemble size representations account for depth in these varied ways, this process is more than a simple readout of V1 spatial frequency detectors.

References

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, *101*, 34–40.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122–131.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.
- Arterberry, M. E. (2008). Infants' sensitivity to the depth cue of height-in-the-picture-plane. *Infancy*, *13*(5), 544–555.
- Barlow, H. B., Blakemore, C., & Pettigrew, D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of physiology*, *193*(2), 327–342.
- Berkeley, G. (1709). *An essay towards a new theory of vision*: Aaron Rhames.
- Bertamini, M., Martinovic, J., & Wuerger, S. M. (2008). Integration of ordinal and metric cues in depth processing. *Journal of Vision*, *8*(2), 10–10.
- Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1160.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*(4), 433–436.
- Burge, J., Peterson, M. A., & Palmer, S. E. (2005). Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision*, *5*(6), 5–5.
- Carlson, V. (1962). Size-constancy judgments and perceptual compromise. *Journal of Experimental Psychology*, *63*(1), 68.
- Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2), 21–21.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393–404.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision research*, *45*(7), 891–900.
- Dakin, S. C., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision research*, *37*(22), 3181–3192.
- Epstein, W., Park, J., & Casey, A. (1961). The current status of the size-distance hypotheses. *Psychological Bulletin*, *58*(6), 491.
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389–1398.
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology-General*, *144*(2), 432–446. doi:<https://doi.org/10.1037/xge0000053>
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception & Psychophysics*, *72*(7), 1825–1838.
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, *18*(5), 855–859.
- Haberman, M., & Ulrich, L. (2019). Precise ensemble face representation given incomplete visual input. *i-Perception*, *10*(1), 2041669518819014.
- He, S., Cavanagh, P., & Intriligator, (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337.
- Heitger, F., Rosenthaler, L., Von Der Heydt, R., Peterhans, E., & Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision research*, *32*(5), 963–981.
- Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *The American journal of psychology*, *54*(1), 21–37.
- Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics*, *71*(2), 375–384.
- Im, H. Y., & Halberda, (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278–286.
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience*, *13*(8), 1020.
- Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20172770.
- Kilpatrick, F., & Ittelson, W. (1953). The size-distance invariance hypothesis. *Psychological review*, *60*(4), 223.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision research*, *35*(3), 389–412.
- Loomis, M. (2001). Looking down is looking up. *Nature*, *414*(6860), 155–156.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4), 6–6.

- 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.
- Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, 7(1-3), 93–106.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, 14(11), 5–5.
- Ono, H., Shimono, K., & Shibuta, K. (1992). Occlusion as a depth cue in the Wheatstone-Panum limiting case. *Perception & Psychophysics*, 51(1), 3–13.
- Ooi, T. L., Wu, B., & He, Z. (2001). Distance determined by the angular declination below the horizon. *Nature*, 414(6860), 197–200.
- Raidvee, A., Toom, M., Averin, K., & Allik, J. (2020). Perception of mean, sums, and areas. *Attention, Perception, & Psychophysics*, 1–12.
- Rensink, R. A., O'Regan, K., & Clark, (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373.
- Ross, H. E., & Plug, C. (1998). The history of size constancy and size illusions.
- Segall, M. H., Campbell, D. T., & Herskovits, M. (1966). *The influence of culture on visual perception*: Bobbs-Merrill Indianapolis.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074.
- Solomon, A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14), 19.
- Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention Ensemble Perception of a Crowd's Gaze. *Psychological Science*, 25(10), 1903–1913.
- Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental science*, 18(4), 556–568.
- 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.
- Watamaniuk, S., & Duchon, A. (1992). The Human Visual-System Averages Speed Information. *Vision research*, 32(5), 931–941.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, 69(1), 105–129.

Open Practices Statement The data for all experiments are on the Open Science Framework (DOI 10.17605/OSF.IO/7U9EN). Experimental and analysis scripts may be furnished upon request. Experiments were not preregistered.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.