Acta Psychologica 142 (2013) 245–250

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

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

Ensemble representations: Effects of set size and item heterogeneity on average size perception

Alexander P. Marchant [a,*], Daniel J. Simons [b], Jan W. de Fockert [a]

[a] Department of Psychology, Goldsmiths, University of London, London, UK
[b] Department of Psychology, University of Illinois at Urbana-Champaign, IL, USA

ARTICLE INFO

Article history:
Received 21 December 2011
Received in revised form 1 November 2012
Accepted 6 November 2012
Available online 30 January 2013

PsycINFO classification:
2323 Visual Perception
2340 Cognitive Processes
2346 Attention

Keywords:
Ensemble representations
Statistical processing
Sampling
Average size
Set size

ABSTRACT

Observers can accurately perceive and evaluate the statistical properties of a set of objects, forming what is now known as an ensemble representation. The accuracy and speed with which people can judge the mean size of a set of objects have led to the proposal that ensemble representations of average size can be computed in parallel when attention is distributed across the display. Consistent with this idea, judgments of mean size show little or no decrement in accuracy when the number of objects in the set increases. However, the lack of a set size effect might result from the regularity of the item sizes used in previous studies. Here, we replicate these previous findings, but show that judgments of mean set size become less accurate when set size increases and the heterogeneity of the item sizes increases. This pattern can be explained by assuming that average size judgments are computed using a limited capacity sampling strategy, and it does not necessitate an ensemble representation computed in parallel across all items in a display.

1. Introduction

The mechanisms that allow us to focus attention on spatial regions, bind features, and recognize objects are limited in capacity (to about four items: Cowan, 2001; Luck & Vogel, 1997). Yet, we can perceive the gist of a scene at a glance, an ability that belies those attention limits. One way to reconcile our proficient perception of the full extent of a scene with the limits of attention is to posit an ability to extract those features necessary for gist perception in parallel across a scene. Scene summary statistics might provide such a source of information.

Observers can extract the preponderant direction of motion of multiple moving dots (Watamaniuk & Duchon, 1992) as well as the average orientation of multiple gratings (e.g., Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) with little effort and seemingly without requiring attention to focus on each item. In these cases, the visual system could extract a statistical summary by pooling across neuronal populations that specifically code for variations in that feature dimension. Intriguingly, people also are able to judge the average size of an array of objects rapidly and accurately, even when they cannot identify the individual items in that array (see Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b; De Fockert & Marchant, 2008; Marchant & De Fockert, 2009). Unlike motion and orientation, there are no absolute size receptors in the early visual system that could serve as a source for such ensemble representations, so the perception of average size would seem to require a different sort of mechanism.[1]

Given the theoretical implications of a mechanism that could extract the sizes of individual objects without focusing attention on those objects, it is essential to test whether performance in these size judgment tasks actually exceeds what could be accomplished via focused attention. In line with a more mundane focused-attention account, many of the extant findings in the size averaging literature are consistent with a strategy of sampling the sizes of several items using focused attention (Myczek & Simons, 2008; Simons & Myczek, 2008); simulations of such a sampling strategy do a reasonable job of approximating the levels of performance exhibited by subjects in studies of average size perception.

However, the fact that sampling strategies approximate human performance does not mean that participants actually use such strategies (Chong, Joo, Emmanouil, & Treisman, 2008). Chong et al. (2008) used stimulus arrays that required different focused attention strategies, intermixed them within a block, and found no cost to average

* Corresponding author at: Department of Psychology, London South Bank University, 103 Borough Road, London SE1 0AA, UK. Tel.: +44 20 7815 5776; fax: +44 20 7815 5775.
E-mail address: alex.marchant@gmail.com (A.P. Marchant).

[1] It may be possible that size is inferred from receptive field size or spatial frequency information. However, spatial frequency information would not be effective for outline circles, and to pool across different receptive field sizes it would be necessary to factor viewing distance into the pooling mechanism.

0001-6918/$ – see front matter
http://dx.doi.org/10.1016/j.actpsy.2012.11.002

size perception compared to a blocked design that did not require strategy switching. Given the lack of strategy switching costs, Chong et al. (2008) concluded that participants were more likely to extract an ensemble size representation using a parallel averaging mechanism, rather than shift among multiple focused attention strategies. Note, though, that participants in that study need not switch among strategies: a single focused attention strategy of averaging the largest and smallest from each side of the display and comparing those approximates the averaging performance achieved in that task (Simons & Myczek, 2008). In order to demonstrate the necessity for a new parallel size averaging mechanism, it is essential to show that the results could not be accomplished via any focused attention strategy.

The claim that performance on average size judgments requires a parallel processing mechanism is based on more than the speed and accuracy of such judgments. Key to these claims is the finding that averaging performance is relatively unaffected by the number of items in the array (e.g., Chong & Treisman, 2003, Exp 1), even when that number exceeds the capacity of focused attention (about four items — Cowan, 2001; Luck & Vogel, 1997). For example, when comparing the average size of an array of circles to the size of a test probe, performance was comparable with array sizes of 4, 8, 12 or 16 circles (Ariely, 2001). Likewise, when subjects judged which of the two sets of circles had the larger mean size, performance was comparable when each set had 8 circles or 16 circles, leading the authors to conclude that "averaging is unaffected by display size" (Chong & Treisman, 2005b, Exp 1, p. 894). The apparent immunity of mean size estimations to the number of set members is key to the idea that ensemble representations might aid perception by extracting the important summary information from complex scenes (Chong, Joo, Emmanouil & Treisman, 2008; Oliva & Torralba, 2007; Treisman, 2006).

Yet, the studies finding no effect of set size on average size perception manipulated set size while strictly maintaining set regularity in terms of the heterogeneity of the item sizes within the set. For example, Ariely (2001) used four set sizes, but only four distinct circle sizes were used in all cases (i.e., set size 12 included three circles of each size). In Chong and Treisman's (2005b) study, only two distinct sizes were used regardless of set size. As a result, the heterogeneity of the set was equal across different set sizes. This regularity means that a sampling strategy will be equally effective with small and large set sizes (Myczek & Simons, 2008).

Two recent studies of average size perception have varied set size with heterogeneous sizes of objects (Corbett & Oriet, 2011; Robitaille & Harris, 2011). Participants viewing an RSVP stream of individual circles of varying sizes successfully judged whether a target circle had the same size as the average of the stream despite being unable to identify an individual member from the stream (Corbett & Oriet, 2011). Although performance is consistent with averaging across heterogeneous items without any cost of increasing set size, the demands on attention that allow averaging across an RSVP task differ considerably from the demands for a single display presented for 1 s or longer (e.g. Chong & Treisman, 2005a, Exps 1, 3 and 4; and the paradigm used in this paper). Simultaneous presentation either requires spatially distributed attention (for parallel processing) or perhaps sequential shifts of attention (for a more serial mechanism). Performance with simultaneous presentations is limited by the speed or spatial capacity of the mechanism performing the averaging (how many objects can be processed at any one time). In contrast, in an RSVP task, attention is focused and the spatial limits on performance are removed—the only remaining limits are temporal. The RSVP study suggests a high-capacity mechanism that is insensitive to heterogeneity, but it does not directly address the distribution of attention to multiple objects in a scene.

One other finding suggests averaging that exceeds the limits of focused attention. Observers viewed a target circle followed by an array of varying numbers of circles of heterogeneous sizes and judged whether the mean of the array was larger or smaller than the target (Robitaille & Harris, 2011). Performance was more accurate for larger set sizes, without a steep increase in response times for larger sets. Although this finding suggests the need for a parallel averaging mechanism that exceeds the limits of focused attention, that conclusion is problematic for several reasons.

First, the difference between the mean size of the set and the size of the target was modified using a staircase procedure to 'keep accuracy within the range of 80–85%' (p. 4). Because this average accuracy was calculated across all conditions, the relative accuracy of the different conditions was not free to vary. For example, if one condition were easy for a participant, the difficulty of all conditions would be increased as a result of their good performance on that condition, thereby inflating (or deflating) the accuracy of the other conditions. Given this lack of independence of performance across conditions, it is difficult to interpret the meaning of accuracy differences across conditions.

Second, it is unclear when the effect was moderated by set size. Much of the set size effect appears to result from a substantial jump in accuracy from set size 2 to set sizes of 4 or more. Performance for set size 2 might be qualitatively different than performance for larger set sizes, resulting from different mechanisms (e.g., Myczek & Simons, 2008). Finally, a focused attention strategy could account for the results: Participants need only determine whether a greater number of the items were larger or smaller than the target circle. This strategy would not produce 100% accuracy (due to skewed sets), but it would achieve an accurate result on the majority of trials without needing to average any items. This calculation, akin to estimating the number of items in a display, is computationally simpler than computing an average of those items.

If average size is computed in parallel across all items in a display, then performance should be relatively unaffected by set size even when the sizes of the items in the display are less homogeneous. If judgments of average size are influenced by item heterogeneity, then claims of a parallel averaging process that automatically extracts the mean size across items would be weakened. At a minimum, such evidence would suggest that not all items contribute equally to judgments of average size (see De Fockert & Marchant, 2008 for evidence consistent with that possibility). It might also be consistent with the idea that average size judgments can be computed by using focused attention to sample a subset of the items in an array (Myczek & Simons, 2008). Subset averaging would predict greater variability in mean estimates as the heterogeneity of the item sizes increases. Across trials, the average absolute deviation of mean judgments from the true mean should be greater with increased heterogeneity because the subsets will include a more variable range of sizes.

In two experiments, we tested whether mean size judgments are affected by the heterogeneity of the items in the array. More precisely, we examined whether the absence of a set size effect in mean size judgments depends on keeping the heterogeneity of the display items constant across set size. Observers viewed sets of circles that were factorially varied in terms of set size and heterogeneity and judged the average size of the set. For each experiment, we used simulations to confirm that subset averaging predicts less precision with more heterogeneous arrays.

2. Experiment 1

Participants viewed a set of circles and then estimated the mean size of the set by adjusting the size of a single circle. Sets contained either 4 or 8 items, comprised of just two distinct sizes (regular sets) or all different sizes (irregular sets). Parallel averaging predicts little or no effect of heterogeneity or set size (Ariely, 2001; Chong & Treisman, 2005b, Exp 1), but subset averaging predicts reduced precision with increased heterogeneity of item sizes (see simulation below).

2.1. Method

2.1.1. Participants

Thirteen undergraduate students (five males; age range: 20–35; mean age: 26) from Goldsmiths, University of London participated in