Acta Psychologica 146 (2014) 7–18

Contents lists available at ScienceDirect

Acta Psychologica

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

Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert

Igor S. Utochkin *, Natalia A. Tiurina

National Research University "Higher School of Economics", Russian Federation

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 23 November 2013
Accepted 29 November 2013
Available online 18 December 2013

PsycINFO classification:
2323 Visual Perception
2340 Cognitive Processes
2346 Attention

Keywords:
Ensemble representation
Visual summary statistics
Size averaging
Parallel processing
Visual sampling

ABSTRACT

In their recent paper, Marchant, Simons, and De Fockert (2013) claimed that the ability to average between multiple items of different sizes is limited by small samples of arbitrarily attended members of a set. This claim is based on a finding that observers are good at representing the average when an ensemble includes only two sizes distributed among all items (regular sets), but their performance gets worse when the number of sizes increases with the number of items (irregular sets). We argue that an important factor not considered by Marchant et al. (2013) is the range of size variation that was much bigger in their irregular sets. We manipulated this factor across our experiments and found almost the same efficiency of averaging for both regular and irregular sets when the range was stabilized. Moreover, highly regular sets consisting only of small and large items (two-peaks distributions) were averaged with greater error than sets with small, large, and intermediate items, suggesting a segmentation threshold determining whether all variable items are perceived as a single ensemble or distinct subsets. Our results demonstrate that averaging can actually be parallel but the visual system has some difficulties with it when some items differ too much from others.

1. Introduction

Numerous studies of the past 10–12 years have demonstrated a striking ability of human observers to rapidly judge the average size of multiple objects scattered over the visual field (Alvarez, 2011; Ariely, 2001; Chong, Joo, Emmanouil, & Treisman, 2008; Chong & Treisman, 2003, 2005b; Robitaille & Harris, 2011, etc.). Size averaging is considered to be a part of the powerful domain of visual processing referred to as ensemble representation that includes computing summary statistics across the variety of perceptual dimensions (see Alvarez, 2011, for review). Ensemble representation is believed to be an efficient tool that permits to encode the properties of lots of items momentarily and thus overcome the severe limitations of visual attention and working memory (Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988).

Many theorists assumed that the processes responsible for representing ensembles are parallel by nature and being provided either by preattention (Chong & Treisman, 2003), or by distributed attention (Alvarez, 2011; Chong & Evans, 2011; Chong & Treisman, 2005a; Treisman, 2006). Two principal findings underlie this assumption. First, averaging performance does not decrease with the set size (Ariely, 2001; Chong & Treisman, 2005b, Experiment 1) or even benefits from increasing the number of items (Chong et al., 2008, Experiment 2; Robitaille & Harris, 2011). Second, the accuracy of judging the average typically remains rather high even while individual representations appear to decay or be practically lost (Alvarez & Oliva, 2009; Ariely, 2001; Haberman & Whitney, 2011; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001).

However, a more precautious view on parallel averaging was suggested by Myczek and Simons (2008; Simons & Myczek, 2008). They argued that the good level of size averaging can be accomplished via a limited-capacity sampling strategy. Such a strategy implies focusing attention on few items and judging the mean based on these items. Allik, Toom, Raidvee, Averin, and Kreegipuu (2013) combined behavioral and simulation data and concluded that the sample size increases with the total set size but rarely exceeds four items (at least sets of no more than eight items they have actually tested). Consistent with this focused attention framework, De Fockert and Marchant (2008) found that average judgments are systematically biased towards the sizes of attended individuals. Definitely, this finding demonstrates that attention is involved in size averaging, but it does not imply necessarily that averaging cannot be carried out in parallel. In the task used by De Fockert and Marchant (2008), observers had to average between all items while locating one particular item, either the largest, or the smallest one in a set. This localization task caused intentional narrowing of the attentional window that could lead to two consequences. First, averaging performance typically gets poorer under narrowed attention (Chong & Treisman, 2005a). Second, attended items could gain

* Corresponding author at: Psychology Dept., Higher School of Economics, 109316, Volgogradsky Prospect, 46-B, Room 321, Moscow, Russian Federation.
E-mail address: isutochkin@inbox.ru (I.S. Utochkin).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.11.012

greater statistical weights when the average was being computed (Pavlovskaya, Bonneh, Soroker, & Hochstein, 2011) that could elicit a corresponding bias. Both these consequences, however, do not exclude the possibility that the rest of the items could be processed concurrently, though their representations could be attenuated (Alvarez, 2011). However, when a task does not require focusing on particular items it is likely that attention is distributed more uniformly among all items yielding more accurate averaging that is necessary for global gist perception.

In their new study, Marchant, Simons, and De Fockert (2013) provided new evidence that averaging can be accomplished by focusing attention on a few items rather than by parallel processing of all items. Reviewing the literature on size averaging Marchant et al. (2013) concluded that the most of the researchers have used highly regular sets including no more than four different item sizes per ensemble. That is, averaging could remain near perfect despite all set size manipulations because there was always the same number of features to be averaged, and this number has never exceeded the limited capacity of attention and working memory. So, in theory, if the observer can focus attention on four differently sized items in such a regular display, he or she probably would be able to average between them without processing other items. Note that to maintain constant averaging efficiency, the observer should be able to always select an ideally representative sample that includes all features and saves their proportions. It is questionable though whether the observer would always succeed in doing this at a brief glance when there are too many items to choose out and individual features cannot be encoded properly. Therefore, relying on the sampling strategy can cause a decrement in averaging performance with the set size even when an ensemble is highly regular. In contrast, parallel averaging does not predict such a decrement because the sample is always exhaustive in that case.

Advocating the sampling mechanism of size averaging, Marchant et al. (2013) conducted two critical experiments for distinguishing between this mechanism and parallel averaging proposed by other authors (e.g., Ariely, 2001; Chong & Treisman, 2003, 2005b; Chong et al., 2008; Robitaille & Harris, 2011). In both experiments, they manipulated set size and regularity of visual ensembles. In their regular condition, all items could have either of two sizes, 0.1° larger or smaller than the mean. In the irregular condition, the number of item sizes increased with the number of items from 4 up to 8 in Experiment 1, or up to 16 in Experiment 2. Every single item had a unique size in that condition. Half of sizes were below the mean and another half of sizes were above the mean (the step of size increment was always 0.1°). Marchant et al. (2013) found in the result that observers were able to keep rather good performance with all set sizes in regular condition that is consistent with previous data (Ariely, 2001; Chong & Treisman, 2005b; Chong et al., 2008; Robitaille & Harris, 2011). However, performance tended to gradually decrease with set size in the irregular condition because the number of features increased as well. This finding provides evidence in favor of using sampling strategy. Obviously, if observers had averaged between all items in parallel their judgments would have been equally accurate regardless of heterogeneity. In contrast, the growing error that actually took place indicates an increasing failure to choose an appropriate sample that would provide a good approximation of the ensemble mean.

However, Marchant et al. (2013) manipulations with regularity affected two stimulus variables simultaneously. The first factor is *heterogeneity*, a number of unique features in a display, which Marchant et al. supposed to be critical for testing their hypothesis. The second variable is the *range of size variation* that indicates the width of the feature distribution among the members of an ensemble. In the regular condition of Marchant et al.'s experiments, observers were always exposed to only two different sizes that were very similar with each other (the difference was always 0.2°) and, most importantly, with the mean. In contrast, irregular sets tended to involve more different increasing and decreasing sizes as the number of items increased. Hence, there was much more dissimilarity between smallest and largest members in irregular displays than in regular ones, and this dissimilarity was growing with set size (for example, it was 1.6° when set size was 16). This implies an alternative to the explanation of the pattern described by Marchant et al. (2013). Perhaps, it is simply more difficult for the visual system to average between dissimilar items.

At least several observations described in the size averaging literature indicate that the range of variation can be related to averaging performance. The first one was documented by Chong and Treisman (2003, Experiment 3) as a minor note. They tested the accuracy of matching average sizes of two ensembles presented to both sides from fixation. Statistical distributions of sizes within an ensemble could be normal, homogeneous, uniform, or two-peaks. Whatever types of distributions were to be matched, the thresholds of mean discrimination were about the same. There was, however, a slightly higher threshold when a two-peaks distributed had been matched to a homogeneous one (but only in naïve, not in trained observers). A core difference between these two types of distributions is that the homogeneous ensembles had only elements representing the exact mean, while two-peaks ensembles had only extremely small and large elements that are the most dissimilar with that mean. More solid evidence for the important role of the range of variation in size averaging was obtained in later studies. Im and Halberda (2013) have shown that increment in size variance of an ensemble causes mean discrimination error to increase as well. Alas, since individual sizes have been chosen randomly from Gaussian distributions in Im and Halberda's study, it is difficult to say whether heterogeneity covaried with variance or not (dissociation between these two variables was not the goal of the study). Fouriezos, Rubenfeld, and Capstick (2008) found that size variance reduces confidence in average judgments. Corbett, Wurnitsch, Schwartz, and Whitney (2012) showed that susceptibility to contrast illusion caused by adaptation to the mean size is reduced by variance, suggesting difficulties in encoding the mean from highly variable sets.

We argue that the range of variation could be a potentially interfering factor in the study by Marchant et al. (2013) that could lead them to conclude about seemingly limited-capacity averaging process. We predict, in contrast, that averaging can in fact be parallel if the range of variation is properly controlled. We tested this prediction in three experiments. In the first experiment, we replicated Marchant et al.'s Experiment 2, the most elaborated of their two experiments, with minor modifications in response registration. In this experiment, the range of size variation increased with set size in the irregular condition. In the second experiment, we equalized the range of size variation between regular and irregular sets and between all set sizes in order to estimate whether it would prevent visual averaging from decay with larger heterogeneity. Finally, the third experiment was aimed at clarifying the effects of local similarity between neighboring sizes on averaging.

2. Experiment 1

Experiment 1 was aimed at replicating the original pattern reported by Marchant et al. (2013) with our own apparatus and stimuli. We tried to reproduce all stimulation as close as possible to their Experiment 2. The only important change was made to response registration system. We replaced the staircase adjustment procedure used by Marchant et al. (2013) by the four-alternative forced choice (4AFC). Marchant et al. allowed their observers to adjust the size of a test circle – pixel by pixel – to achieve an apparently average size. A starting test size could be obviously far from the mean (below the smallest or above the largest member of a presented set), and observers had to make a lot of steps to achieve a plausible range where the mean could be. This procedure has an advantage because it catches even tiny increments and decrements in observers' estimates in near analogous mode. On the other hand, as adjustment takes time, a memory trace of the