Vision Research 83 (2013) 25–39

Contents lists available at SciVerse ScienceDirect

Vision Research

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

An almost general theory of mean size perception

Jüri Allik [a,b,*], Mai Toom [a], Aire Raidvee [a], Kristiina Averin [a], Kairi Kreegipuu [a]

[a] Department of Psychology, University of Tartu, Estonia
[b] Estonian Academy of Sciences, Estonia

ARTICLE INFO

Article history:
Received 14 September 2012
Received in revised form 10 February 2013
Available online 13 March 2013

Keywords:
Mean size perception
Ensemble characteristics
Compulsory pooling
Associative law of addition
Internal noise
Capacity limitation
Selection

ABSTRACT

A general explanation for the observer's ability to judge the mean size of simple geometrical figures, such as circles, was advanced. Results indicated that, contrary to what would be predicted by statistical averaging, the precision of mean size perception decreases with the number of judged elements. Since mean size discrimination was insensitive to how total size differences were distributed among individual elements, this suggests that the observer has a limited cognitive access to the size of individual elements pooled together in a compulsory manner before size information reaches awareness. Confirming the associative law of addition means, observers are indeed sensitive to the mean, not the sizes of individual elements. All existing data can be explained by an almost general theory, namely, the Noise and Selection (N&S) Theory, formulated in exact quantitative terms, implementing two familiar psychophysical principles: the size of an element cannot be measured with absolute accuracy and only a limited number of elements can be taken into account in the computation of the average size. It was concluded that the computation of ensemble characteristics is not necessarily a tool for surpassing the capacity limitations of perceptual processing.

1. Introduction

Knowing how much Sir Francis Galton was obsessed with counting, it is not surprising that, when he visited a livestock fair in Plymouth in 1906, he collected all 787 ballots on which participants of a weight-judging competition marked their guesses about the weight of a particularly fat ox, after it was slaughtered and dressed. Galton published a note in *Nature,* showing that, if the democratic principle "one vote, one value" was applied, the mean of all 787 guesses was only a pound off the actual weight of 1,198 lb, which was closer than the prediction made by any single voter (e.g. Galton, 1907). James Surowiecki documented in his well-received book this and many similar examples demonstrating the wisdom of the crowd (Surowiecki, 2004). These demonstrations served to rehabilitate, partly at least, the image of the crowd, which was usually portrayed, at least since the publication of *The crowd* (Le Bon, 1896), a famous book by Gustave Le Bon of France, as sentimental, hysterical, and with an intellect equal to that of a small child or a "savage."

Later studies, usually under the heading "intuitive statistician," demonstrated that not only a crowd but also individual observers can make, occasionally at least, remarkably accurate statistical judgments (Peterson & Beach, 1967). Please have a brief look at the following list of two-digit numbers and try to guess, immediately after reading, them what their arithmetic mean is. Here they are: 54, 25, 79, 39, and 83. What was your guess? If your guess is no more than few numbers off 56, then you may be proud of your statistical intuition. Many studies have shown that human observers give estimates surprisingly close to the actual mean, even if the amount of numbers approaches 20 or more and there is no time for elaborate mental arithmetic (Anderson, 1967; Smith & Price, 2010). A reasonable explanation is that we all have an intuitive number sense which allows us to provide an approximate but reasonably accurate solution to different statistical tasks (Dehaene, 2011; Dehaene, Dehaene-Lambertz, & Cohen, 1998).

After replacing numbers with geometric figures, it was also noticed that people can make fairly precise judgments about the average size of several geometric objects, typically lines or circles (Ariely, 2001; Chong & Treisman, 2003, 2005b; Fouriezos, Rubenfeld, & Capstick, 2008; Miller & Sheldon, 1969; Spencer, 1961, 1963). These studies were mainly inspired by an observation that the reduction of a set of similar items to a prototypical mean helps to economize on the limited capacity of the visual system by replacing multiple representations of individual elements with a statistical summary characterizing the set as a whole (Ariely, 2001; Chong & Treisman, 2003; Chong et al., 2008). These studies have also demonstrated that the mean size of a group of geometric figures can be judged almost as precisely as the length of a single line or the size of a circle. Applying statistical rules to the perception of mean size, one could expect that precision will increase

* Corresponding author. Address: Department of Psychology, University of Tartu, Tiigi 78, Tartu 50410, Estonia. Fax: +372 7376152.
E-mail address: juri.allik@ut.ee (J. Allik).

0042-6989/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2013.02.018

with the number of judged elements. Since the precision of the sample mean improves with the square root of the sample size ($\sqrt{N}$), it is also expected that mean size judgment would improve with the square root of the elements to be judged (e.g. Fouriezos, Rubenfeld, & Capstick, 2008). In fact, the precision of mean size judgment is typically found to be independent of the number of judged elements (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005b; Fouriezos, Rubenfeld, & Capstick, 2008; Spencer, 1961, 1963), which indicates that the human observer is not fully able to use the advantages of statistical aggregation. However, some researchers have seen the irrelevance of the total number of elements as evidence that the process of perceptual averaging is carried out automatically, presumably by an array of parallel "computers" (Alvarez, 2011; Ariely, 2001, 2008; Chong & Treisman, 2005b; Chong et al., 2008). The idea of involuntary and massively parallel processing of mean size was substantially strengthened by the finding that observers can fairly accurately report the average size of an array of geometric objects, even when they cannot recall the size of individual elements in the array (Ariely, 2001; Chong & Treisman, 2003, 2005a; Corbett & Oriet, 2011). Since mean size can be computed without explicit knowledge about individual elements, this has been taken as additional evidence in favor of automated and parallel computing.

Nevertheless, in spite of considerable interest and an increasing number of studies, the main properties of parallel and effortless computation of mean size remain poorly understood. Even one of the central claims that mean size can be computed outside of focused attention needs to be viewed with reservation since it was demonstrated that all published evidence can be explained, in principle at least, through various focused-attention strategies, without invoking a special mechanism for average size perception (Myczek & Simons, 2008; Simons & Myczek, 2008; Solomon, Morgan, & Chubb, 2011). Indeed, all advocates of effortless and massively parallel computation assume, more or less explicitly, that all or at least nearly all exposed elements are taken into account in the judgment of the mean size. This premise, however, is very difficult to maintain, after many convincing demonstrations that the observer is at times virtually blind to a substantial amount of information. Several well-known phenomena such as the invisible gorilla (Chabris & Simons, 2010; Simons & Chabris, 1999) and inattentional (Simons, 2000) and change blindness (Simons & Rensink, 2005) suggest that the proposition about the use of all available information is unrealistic. Many studies in the framework of ideal observer analysis have also shown that, in numerous situations, the observer is able to use only a fraction of available information (Burgess et al., 1981; Raidvee et al., 2011; Rose, 1948). In the most extreme example, Myczek and Simons (2008) demonstrated that an observed precision of about 4–7% in the judgment of mean size can be explained by assuming that only 2–3 elements are enough to achieve accuracy in the mean size judgment observed.

Another reason why there is not a sufficiently general theory is that the previous research on the perception of mean size has identified several vital but still facultative properties in the computation of average size. It was observed, for instance, how concentration of attention (Alvarez & Oliva, 2008, 2009; Ariely, 2001), different visual cues (Alvarez, 2011; Chong & Treisman, 2005a), rapid temporal presentation (Corbett & Oriet, 2011; Joo et al., 2009), minimally required exposure time (Whiting & Oriet, 2011), resistance to object substitution masking (Jacoby, Kamke, & Mattingley, 2012) and previous adaptation (Corbett et al., 2012) affect the ability to estimate mean size. Surprisingly little attention has been paid to the defining properties of the statistical averaging process itself. For example, it is well known that the order in which addends are summed does not change the end result. Similarly, the grouping of added numbers does not affect the sum (the associative law). If the observer's task, for example, is to discriminate the mean size of four circles in comparison to a reference, then it does not matter whether we add, for instance, 4 size units to the diameter of only one of them or we add one size unit to the diameters of all four circles. Intuitively, it is more likely that the human observer can more easily notice an outlier which is 4 size units larger than the reference size, rather than four small increments of 1 size unit added to each of the four circles. Albeit counterintuitive, any theory insisting that the perceptual system is capable of computing mean size must confront the challenge of showing that these two cases result in an identical perceptual outcome. Grouping these 4, or any other number of units, into different packages does not affect the sum nor consequently, in theory, the perception of the mean size. As far as we know, this very easily falsifiable prediction has never been tested before.

The series of studies reported in this paper were conducted with the goal of carrying out a systematic study of mean size perception. In addition to testing some of the fundamental properties of statistical averaging, we will propose a very simple and almost universal explanation for a large range of facts related to the perception of mean size. By theory, we understand a formal system expressed in exact mathematical terms which is able to reproduce all observer response functions with the required accuracy. The proposed theoretical explanation should be sufficiently general so that it depends on only some universal stimulus attributes (number of elements, their sizes, etc.), not particular details (color of elements, their exposure time, sharing of attention between two tasks, etc.) which could vary from one situation to another, but have only marginal effects on mean size judgments. The proposed model is also expected to contain a minimal number of free (to be determined) parameters, all of which are anticipated to have a very transparent psychological interpretation.

One of the two basic assumptions on which the proposed Noise and Selection (N&S) Theory is based on, is that neither the size of a single object nor the mean size of a set of objects can be measured with absolute precision. The size of every single element must be measured and transformed into its subjective representation. This process of transformation is inherently noisy and the measured value varies from one trial to another, even if the physical size remains constant. This means that the perceived size is represented on a psychological continuum, with some positional error, what Thurstone called "discriminal dispersion" (Thurstone, 1927a, 1927b). This dispersion is obviously not a constant, but rather demonstrates a systematic relationship with some stimulus parameters. One of the very first attempts in the history of experimental psychology established that the just noticeable difference (JND) between two sizes increases approximately in proportion to the total size of these objects (for a review see Wolfe, 1923). The interest of the founders of experimental psychology as well as the current researchers in this quantitative relationship, usually known as Weber's "law," was mainly motivated by an understanding that, from an error function of discrimination, it is possible to recover the metrics that the observer is using to make decisions about spatial intervals, or any other visual attribute (Dzhafarov & Colonius, 2011). The existence of non-zero discriminal dispersion means that, even if the number of processed elements is well within capacity limitation, their size cannot be judged without error (Im & Halberda, 2012; Solomon, Morgan, & Chubb, 2011).

On the other hand, we need to assume that, if the number of judged elements exceeds the capacity of attention or working memory, then it is very unlikely that all available information can be processed. In order to explain experimental data, it is necessary to assume that the observer's decisions about a large number of elements may be actually based on a substantially smaller subsample of these. When the number of judged elements exceeds capacity limitation, a sizable number of elements is simply ignored and the presence of the elements in the stimulus has no effect on