University of Tartu

Institute of Psychology

Kristiina Averin

PERCEPTION OF THE MEAN AND SUM SIZE OF GEOMETRIC FORMS

Master's thesis

Supervisors: Jüri Allik, *PhD*

Aire Raidvee, *PhD*

Running head: Mean and sum size perception

Tartu 2013

Contents

Abstract

Human efficacy in mean and sum size estimation was tested in this thesis. Kahneman (2011) proposed mean and sum size of geometric figures to be estimated by different systems – System1 and System 2. Effortless and automatic System 1 allows estimating mean size with considerable accuracy. Sum size, which requires multiplication of the means, however, can be only computed by a more elaborate higher order system, System 2. Two experiments, sharing the test elements, but with different reference and instruction, were conducted to test Kahneman's proposal. In the first experiment the observers were asked to estimate mean size of a set of elements; in the second, the task was to estimate the sum size of the same elements. We expected to see great differences in the accuracy of size and sum discrimination if the underlying operations used in these tasks were different. The results show sudden drop in the accuracy if participants were required to estimate the sum size instead of mean size. Instead of assuming multiplication in sum size estimation, we proposed a model, where all the elements are set side-by-side, following an imaginary line, with the sum distance occupied by the adjoining elements being estimated instead. Accuracy is lowered in the sum size discrimination task by the measurement error of single elements, which is likely to be increased by the additional requirement – mental transposition of the elements – that one could estimate the required property. In addition, we could see that the mean size of a set of similar elements can be estimated only by using a subset of 2 –3 of all elements. Therefore, accuracy in the sum size estimation task can be reduced not only by the transposition need, but also by the requirement to use all the elements for creating an estimate.

Summary in Estonian

Geomeetriliste objektide keskmise ja summaarse suuruse tajumine

Käesolevas uurimuses testiti inimese efektiivsust keskmise ja summaarse suuruse arvutamisel. Kahneman (2011) oletas, et geomeetriliste kujundite summa ja keskmise arvutamiseks kasutatakse erinevaid süsteeme. Nendeks on Süsteem 1 ja Süsteem 2. Automaatne ja pingutust mitte-nõudev Süsteem 1 võimaldab meil võrdlemisi täpselt hinnata sarnaste objektide keskmist suurust, samal ajal kui summaarse suuruse hindamiseks tuleb kasutada keerukamat kõrgema tasandi süsteemi ehk Süsteemi 2. Viimane tuleneb Kahneman’i oletuste kohasel nõudmisest, et summa hindamiseks korrutatakse elementide keskmise suuruse hinnangut elementide arvuga. Viimase oletuse kontrollimiseks korraldati kaks katset, kus kasutati samu test-elemente, kuid muudeti etaloni ja instruktsiooni. Ühel juhul lasti osalejatel hinnata elementide keskmist suurust, teisel juhul samade elementide summaarset suurust. Oletasime, et kui neid kahte ülesannet tehakse erinevate operatsioonide abil, siis on ka täpsus nende ülesannete sooritamises väga erinev. Selgus, et kui ülesandeks oli keskmise suuruse asemel hinnata summaarset suurust, muutusid tulemused märkimisväärselt ebatäpsemaks. Samas, selle asemel, et eeldada summa hindamise protsessis korrutamistehet, esitame mudeli, mille kohaselt selleks, et hinnata summaarset suurust, seatakse mentaalsel tasandil kõik elemendid üksteise kõrvale ritta ühele kujuteldavale joonele ning seejärel hinnatakse nende kõrvutiolevate elementide alla jäävat vahemaad. Täpsus väheneb summa hindamisel antud mudeli järgi seetõttu, et elementide mõttelisel überpaigutamisel iga üksikelemendi mõõtmisel tehtav viga suureneb. Samuti nägime seda, et sarnaste elementide keskmist suurust saab hinnata üksnes 2 – 3 elemendi põhjal. Seega võib täpsus summaarse suuruse hindamisl väheneda lisaks mõttelise ümberpaigutamise nõudele ka seetõttu, et summa hindamine eelkirjeldatud moel eeldab kõikide elementide kasutamist.

The visual system can be understood, besides other things, as an intuitive statistician, since numerous evidence indicate that observers are able to encode and represent ensemble characteristics, computed from multiple individual measures and combined across space and time (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003). An elevated interest towards ensemble characteristics is mainly motivated by a consideration that statistical representation helps to economize on the limited capacity of the visual system. Rather than preserving all the detailed information in a scene, the visual system can abstract the statistical properties and then fill them in at a retrieval using the stored statistics (Chong & Treisman, 2003). Although there were numerous claims that the visual system can effortlessly compute and represent the mean size of a set of similar geometrical objects, typically lines or circles (Alvarez, 2011; Ariely, 2001; Chong, Joo, Emmanouil, & Treisman, 2008; Chong & Treisman, 2005; Joo, Shin, Chong, & Blake, 2009; Miller & Sheldon, 1969; Myczek & Simons, 2008; Solomon, Morgan, & Chubb, 2011; Spencer, 1961, 1963), there was no proof that the size aggregation obeys the axioms of arithmetic addition (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013). In this lastly mentioned study, it was shown that the representation of the mean size is indeed indifferent, whether we add, for instance, 4 size units to the diameter of only one of four test circles presented on the display, 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 test circles. However, the results show, in a good harmony with the associative law, that these two cases result in an identical perceptual outcome, which indicates that the visual system is really insensitive to the grouping of increments being tuned only to their mean size (Allik et al., 2013).

The arithmetic mean is only one ensemble characteristics in a long list of potentially available statistics. The harmonic mean, sum, and standard deviation are only few of many other statistics that came to the mind. Although the mean is conventionally defined as the sum of measures divided by the number of measures, Kahneman (2011, pp. 92-93) came out with an intriguing idea that the average size of the geometric figures can be judged with a considerable accuracy, but the sum size of the same figures cannot. In a sharp contrast to the mean, visual system performs very poorly when sum of the same geometric figures is asked to be judged. According to Kahneman's idea, mean length or size of a collection of nearly identical geometric figures can be computed by the System 1, which is evolutionarily old system producing rapid, parallel and automatic processes, where only the final product is accessible to a cognitive awareness. On the other hand, the System 2 is evolutionarily recent which performs the slower, sequential, and deliberate thinking (Kahneman, 2011). Kahneman believes that the task to estimate the total size activates the System 2, which will laboriously estimate the average, count the number of objects and multiply average size by the number of estimated objects (p. 93). The proposal that sum is derived from mean value is by all means unorthodox. Every technical definition of the “mean” presupposes summation: adding up values and then dividing by the number of values. Unfortunately, Kahneman did not provide any clue as to how it is possible to compute mean without computing sum in the first place. He also leaves us with his imaginative intuition presenting no solid empirical evidences for his proposal that visual system deals well with averages but poorly with sums (Kahneman, 2011, p. 93).

If we understood Kahneman (2011) correctly, he seems to think that it is primarily need of multiplication which slows the System 2 down. There were heated debates in the history of psychophysics whether or not the human observer is capable of estimating sensory ratios. Although Stevens (1975) promoted the direct scaling methods, it was a shared agreement that the fractionation methods are more reliable and accurate in the construction of psychophysical scales (Torgerson, 1958). The logic of the fractionation methods assumes that a subject is capable of reporting or producing the predetermined magnitude of sensory ratios. For example, if a subject is presented with two stimuli, it is presumed that she or he is able to report, with a required precision and consistency, the exact ratios between these two stimuli on their designated attribute. However, human ability to estimate or produce exact sensory ratios was questioned by the results showing that there is no substantial difference whether the subject estimates stimulus ratios or if she or he is asked to divide an interval into two subjectively equal sections (Garner, 1954). Based on these and other results which were collected later, Torgerson (1961) formulated a principle which is known as the Torgerson’s conjecture: the human observer is not able to distinguish between sensory ratios and sensory differences (Birnbaum & Veit, 1974). The Torgerson’s conjecture was both rejected (Luce, 2012) and confirmed (Masin, 2013) by later studies. However, all these debates were based on a good faith that participant is able to follow instruction accurately. If she or he is told to produce or judge one magnitude which is two or any other integer times larger (or smaller) than some reference magnitude, then she or he does it. There are surprisingly little attempts to demonstrate that subjects can indeed follow instruction and multiply or fractionate sensory magnitudes by an exact factor n, not a value that might be slightly off of it. Compared to these debates, the Kahneman's proposal is even more radical: the visual system deals poorly with sums because multiplication is simply beyond its scope and for this reason the task should be passed to the slow thinking system (Kahneman, 2011).

The main idea of the current study is extremely simple. In almost all studies of statistical representation participants were asked to judge the mean size of similar geometric figures which varied in their size. For example, the left panel in Figure 1 shows three randomly positioned circles with unequal diameter. A typical task is to estimate whether the mean size of these three circles is smaller or larger than the size of a reference circle shown on the right side (solid ring). If the Kahneman's intuition is correct, together with numerous previous experimental studies, then this task can be solved with a remarkable accuracy. Indeed, many studies have shown that the mean size can be estimated with the precision of 4-7% from the size to be judged (Myczek & Simons, 2008). However, if the task is only slightly modified with a request to make judgement about the summary size – sum of the three diameters – then it is expected to be an almost impossible mission that can be accomplished with a considerable error if at all. Everyone can test Kahneman's intuition by assessing how difficult or easy is to compare the sum of diameters of these three circles with the size of the reference circle (broken-line circle on the right).

Unfortunately, we are not aware of any attempts to set up an experiment, in which the ability to estimate sum was directly and systematically compared with the ability to judge the mean size. In order to eliminate this gap in our knowledge, the following experiments were planned as a systematic comparison of the observer's performance in estimating the mean and sum size of circles. Provided that the Kahneman's conjecture is correct, we are expecting to obtain much higher precision in the judgement of the mean size but much more impoverished performance in the judgement of the sum size. The expected gap in performance can be explained by differences in operations that are required to judge the mean and the sum size of the figures. If for the mean size it is sufficient to compare the size of each test circle with the size of a reference circle simultaneously present on the display and afterwards accumulate these perceived differences then for the summary size it is necessary to measure the spatial interval covered by circles arranged on a line side-by-side. It is expected that this might be quite difficult task for the visual system. It is relatively easy to compare test objects with the visible reference which has approximately the same size, but it is considerably more difficult to imagine or somehow compute the length of the summary of spatial interval covered by test objects' diameter when they are placed side-by-side.

Methods

Participants. Five participants with self-reported normal or corrected to normal vision participated in both experiments. All observers but one had extensive prior experience with various vision perception experiments.

Apparatus. Stimuli were presented on various LCD monitors that function at least 1,600 × 900 resolutions. In order to compensate for possible variations in screen sizes participants used, the program adjusted stimulus resolution and calculated recommended viewing distance for every given screen size. The adjustments in viewing distance were made to assure that one pixel would subtend to 2 minutes of arc for every participant. Displays contained two adjacent dark circular gray background panels, one for test elements and one for reference element. Each of the background panel was approximately 16.3° in diameter and was presented on a black background. The panels were presented on the left and right side from the central fixation mark with a gap of 44 pixels between them. Experiments were written and run under a program written in MATLAB (The MathWorks, Inc.) using Cogent 2000, developed by the Cogent 2000 team at the FIL and the ICN, and Cogent Graphics, developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience.

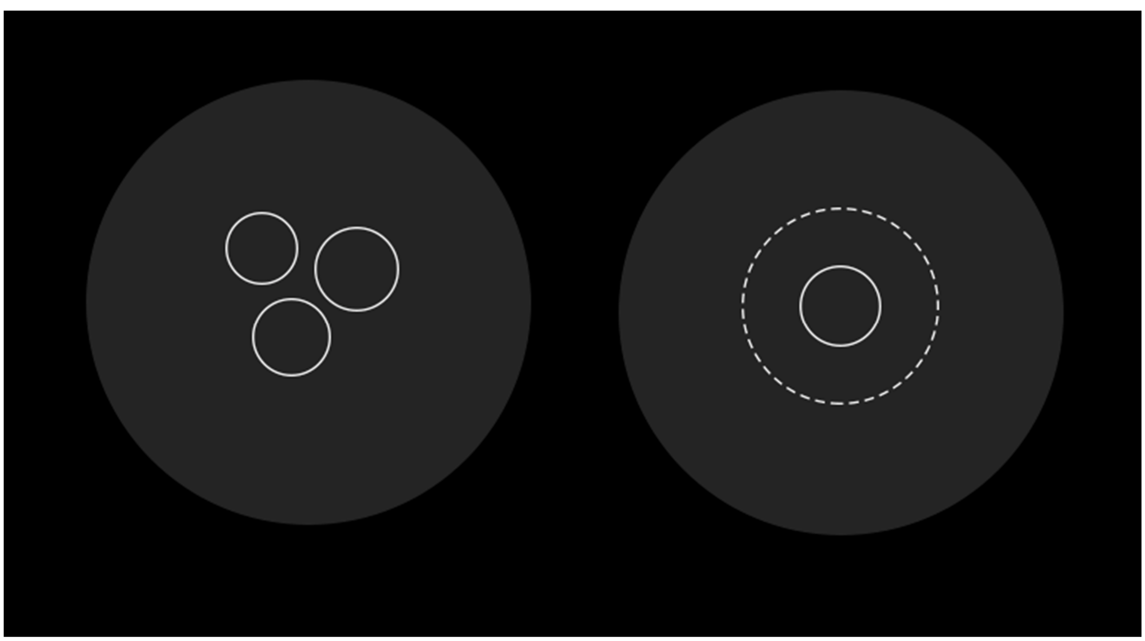

Figure 1. The task was to estimate the mean or sum of three test circles on the left compared with a left right circle corresponding either to the mean (solid line) or sum (broken line) of these three test circles.

Procedure

There were two types of experiments differing from each other by the instruction.

Mean size experiment. Test stimuli for the mean size estimation task consisted of a set of one, two, three, or seven (N = 1, 2, 3 or 7) randomly positioned spatially not overlapping white unfilled circles of various sizes. Background panel, the left or right, was chosen randomly before each trial. The opposite stimulus area belonged to the reference, which size was ether smaller or larger than the mean size of the test elements.

The total diameter of test circles in each trial was 11.3° of visual angle. The base sizes for the reference and test circles in a base-set (set of elements equal in size to which increases and decreases were later added) were calculated by dividing the diameter in single element condition by the number of elements in current condition. Thus, diameter of the reference stimuli in the single element condition was two times longer than the diameter in a condition with two elements and seven times longer than in a condition with seven elements.

In each trial, mean size of the test elements was set to differ from reference by increasing or decreasing the mean size of base-set elements in pixels by a variable delta *(Δd)*, which acquired values -12, -8, -4, -2, 2, 4, 8 or 12 pixels whereas the diameter of every single element was allowed to acquire values in range [$300/N \times 0.95$, $300/N \times 1.05$] (a schematic view of a stimulus is given in Figure 1). Location of each element on the test panel was random. However, to assure that the elements would not overlap or cross a border of a panel, inhibitory area was set around each circle and on the panel borders. In one-element stimuli, the element was always presented at the center of the panel.

Each trial started with 1s presentation of two background panels, one containing test circles and the other containing a reference circle. Value of delta *(Δd)* was selected randomly for each trial. Participant's task in the first experiment was to estimate mean size of the test elements and indicate by corresponding mouse click whether right or left panel had greater average element size. Note that in case of one element test-set, mean size of both, reference and test-set, were represented by the size of the single element and thus the task in this condition fell back to ordinary size comparison task. After response, auditory feedback about the correctness of the answer was given. In case of correct answer a sound with high tone was played and in case of incorrect answer a sound with low tone was played. All the answers were recorded to a text file. Minimum 54 trials were completed by every participant for each condition with each number N of elements and delta *(Δd)*.

Summary size experiment. Test stimulus elements in the summation experiment remained the same as in the mean size experiment. Unlike in the first experiment, the reference in the second experiment remained the same 11.3° throughout the trials for each condition and number of elements in the test stimulus. Total diameter of the base-set elements did not depend on the number of elements.

Total length of circle diameters in the test-set was varied similarly to the mean size estimation experiment. Sum-size of the base-set elements was increased by Δd = -12, -8, -4, -2, 2, 4, 8 or 12 pixels in one element condition and Δd = -36, -24, -12, -6, 6, 12, 24 or 36 pixels for the test-set with more than one circle. Background areas for reference and test stimuli were randomly chosen for each trial. Element locations were selected and inhibitory areas applied similarly to the mean size experiment.

Again, each trial started with presentation of a reference circle and test circles for which delta (Δd) was randomly selected. Similarly to the mean size experiment, reference was again a single circle which represented the sum size of the reference throughout all the trials. Thus, as in the previous experiment, in one element condition the task of estimating sum sizes of the elements was essentially comparing sizes of two elements. Again, auditory feedback about correct and incorrect responses was given. For each condition minimum 54 repetitions were completed and all the answers were saved to a text file for further analyses.

Results

Psychometric curves for mean size discrimination task are given in Figure 2. The columns in the figure correspond to the number of elements in the test panel (N=1, 2, 3, 7). The rows of the panels correspond to the four observers (JA, KA, MT and AR). Probabilities of answering that the mean size of the test circles is greater than the reference circle (vertical axis) are given as a function of the mean size differences in pixels (Δd). Empirical data was approximated by cumulative normal distribution with best fitting mean (μ) and standard deviation (σ), where mean (μ) marks the value of delta (Δd) for which probability of giving the answer that the mean size of the test circles is greater than the reference circle is 0.5. Standard deviation (σ) marks the slope of the psychometric function and the difference in pixels, in which case the observer is able to notice the difference in 84.1% of cases. The greater is the value of σ, the gentler is the slope and the larger is the just notable difference (JND) for the current condition.

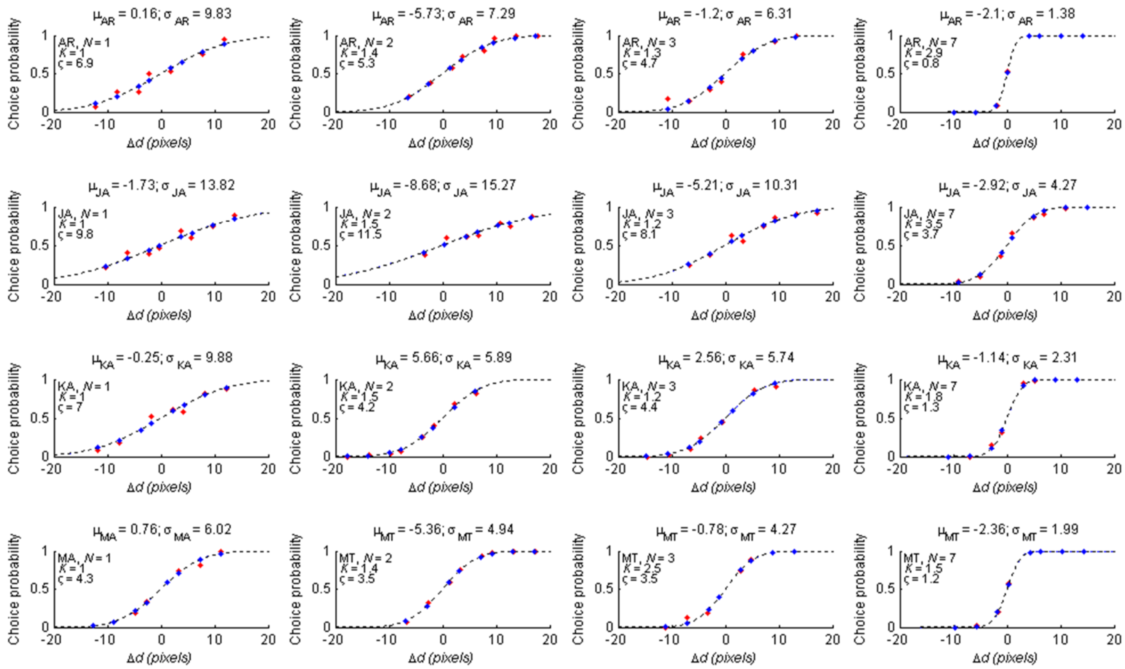

Figure 2. Curves for mean size discrimination task. Red dots represent empirical probabilities that mean size of the test-set elements is answered to be greater than the reference circle for given delta (*Δd)*. Blue dots represent theoretical probabilities received by simulation for empirical data. Dashed lines represent curves of cumulative normal distribution to which both, empirical and theoretical data, were approximated.

In mean size discrimination tasks we can see a trend that the accuracy of mean size discrimination (σ) varies with the number of elements (Figure 4). For all of the participants JND was the lowest for the condition where the mean size of seven elements (*N*=7) was to be estimated. Across participants, there were significant difference between standard deviations of one and seven element condition, and two and seven element condition [$F(3, 15)=5.4$, $p=.014$]. The results are not surprising since the element sets were constructed in a way that the diameter of the base element in case of one element was divided by *N*. Thus, the larger was *N*, the smaller was each element in the set. As we know by Weber's Law, the smaller is the magnitude of the property measured, the smaller is the increase or decrease needed for creating JND between two elements. Moreover, Allik et al. (2013) found that the number of elements is not determinative for the accuracy of mean size discrimination. Myczek and Simons (2008) and Allik et al. (2013) showed that in estimating mean size of similar elements not all elements in the display are necessarily taken into account for making the final decision with the accuracy seen in the empirical data. Allik et al. (2013) proposed a model which assumed that human observers measure elements with an unavoidable random error caused by Thurstonian internal noise. Moreover, if element number *N* in a set of elements exceeds capacity limit of human visual system, a subset *K* of all the elements is used for making the decision about mean size of the whole set of elements.

In their paper Allik et al. (2013) used data simulation for producing a theoretical model which fits empirical data with a remarkable accuracy. They used Noise and Selection (N&S) model where the mean size of *N* was estimated by a subset of elements *K*. In addition, normally distributed random error (ς – final sigma) was applied to each measurement of a test-set element. These results suggested that the number of elements, taken into account by an observer while doing mean size discrimination task, is around four. The measurement error varied from ς=4.1 to ς=11.6 between the subjects. (Allik et al., 3013)

We applied N&S model to the current data. Since in our experiment, the sizes of the elements varied together with the number of test-set elements *N* in different conditions, we expected that the random measurement error may vary between element sizes. Thus, both, *K* and ς were free to vary during the data simulation. However, in case of one element conditions, we also assumed that *K*=*N*, and thus the accuracy in responses is affected only by internal noise (ς). First, the data was simulated under an assumption that in each trial a subset of test-set elements were randomly chosen. Normally distributed random error was added to each element in the subset. Thereafter mean size of the subset of elements was computed and then compared to the reference element. If the average size of the elements was greater than the size of the reference, answer “mean of the test-set is greater” was chosen.

However, the closest fit to the empirical data was achieved if the random error was also added to the measurement of the reference element. Such addition was driven from the fact that the size of reference circle was different for trials with different *N*. Therefore, comparing to the experiment where only reference in one size is used throughout one experimental set up (as in Allik et al., 2013), we assumed that in our experiment, learning effect was less likely to occur. Since it is common for observers to be somewhat biased in their answers and prefer one answer over the other, answering bias was compensated for each subject by shifting means of theoretical data towards the means of empirical data for achieving better fit.

Probabilities of the last described data simulation with *K* and ς values, providing the closest fit to the data, are given with blue dots in Figure 2. As in the previously described study, *K* could acquire non-integer values since the number of elements (*K*) used by the observer for making the decision could vary from trial to trial and thus by averaging over the values of *K* in different trials, the final *K* value for each participant for each value of *N* was achieved. The fit (correlation between observed and predicted data) for all the models was at least r = 0.968. The results are in accordance with previous results. For two element condition the model gave a prediction according to which the observers used in average less than two (starting from K=1.4 to K=1.5) elements. For three and seven element conditions again the best fit was achieved if about two or three elements were expected to be taken into account (K=1.2 to K=2.5 and K=1.5 to K=3.5 respectively). Thus, it seems that the number of elements used for making a decision about mean size remains around the same K value for all N. On the other hand, measurement error varied with some minor deviations together with N (see Figure 2) and thus with the size of reference element. For all the observers measurement error was the lowest in seven element condition and for three participants out of four the highest for the one element condition. It proves again that the number of elements in the test-set do not have a substantial impact on the discrimination accuracy while decrease in the size of the test-set elements entails decrease in measurement error.

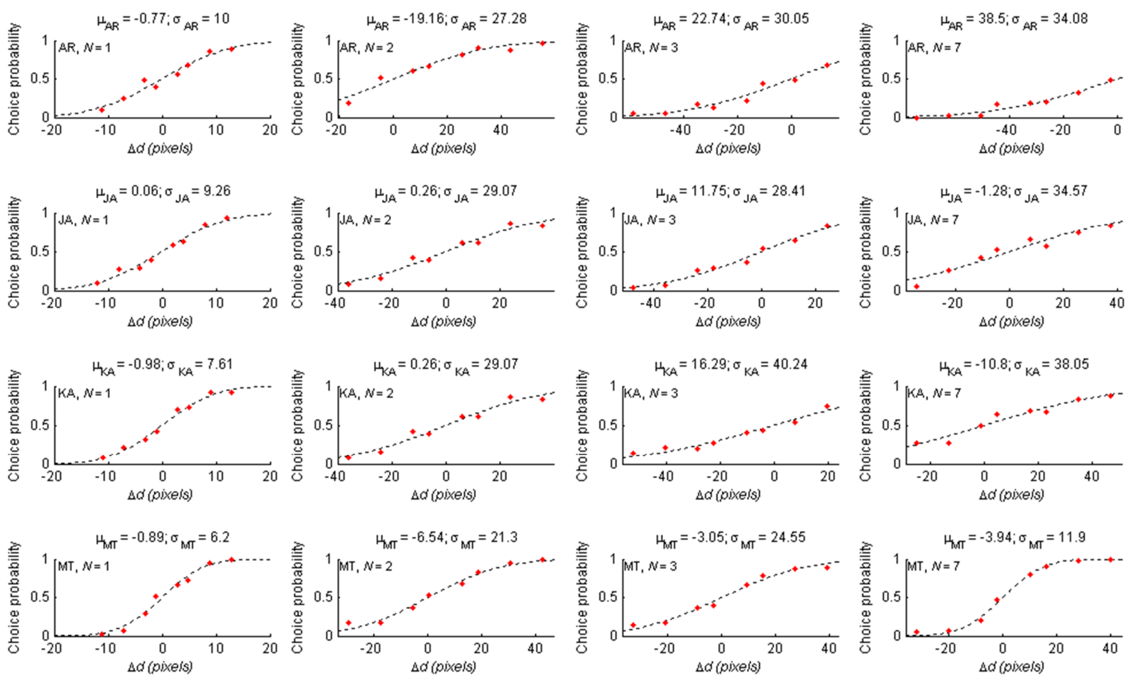

Figure 3. Curves for summary size discrimination task. Red dots represent empirical probabilities of giving the answer that the test-set elements have greater sum size than the reference circle. The dashed line represents curve of cumulative normal distribution to which the empirical data was approximated.

Curves for sum size discrimination task are given in Figure 3. Similarly to the mean size discrimination task, columns correspond to the number of elements and rows correspond to the observers. Again, delta *(Δd)* is given on the horizontal axis and probabilities for answering that the sum size of the test elements is greater than the reference element on the vertical axis. Like in the mean size estimation task, cumulative normal distribution was fitted to the data.

As expected, discrimination accuracy in one element condition did not differ substantially between the experiments (see Figure 4) [$F(1, 7)=0.36$, $p=0.46$]. Since these two conditions $N=1$ for the mean and sum size decision were identical, it could be treated as a test for the explicability. For the other conditions ($N>1$), the results in summary size estimation experiment do not follow the same pattern with the results of mean size estimation task. While increase in the number of elements in the mean size discrimination task led to a slight increase in accuracy, in the sum size discrimination task estimation the accuracy was substantially impaired by the increase of the number of elements. Across the participants standard deviation (σ) was significantly higher in one test-set element ($N=1$) condition than in conditions with more elements (N= 2, 3, 7) [$F(3, 15)=7.95$, $p=.004$].

Figure 4 plots the standard deviation in the discrimination performance for the two tasks, mean and sum size discrimination, as a function of the number of elements. It is immediately clear that – except physically identical conditions ($N=1$) – the discrimination accuracy is radically different for mean and sum size task: it is considerably more inaccurate to tell summary than mean size of a set of circles. Despite the fact that delta *(Δd)* in two, three and seven item conditions is three times larger for the summary size discrimination task, the accuracy in discriminating sum of elements is respectively about five (σ=6.36 for mean size and σ =30.82 for sum size) and ten (σ=2.33 for means size and σ=29.65 for sum size) times lower than the accuracy in mean size discrimination task. Thus, all four observers were able to determine the mean size with a remarkable precision while accuracy of summary size estimation of exactly the same elements was many times worse.

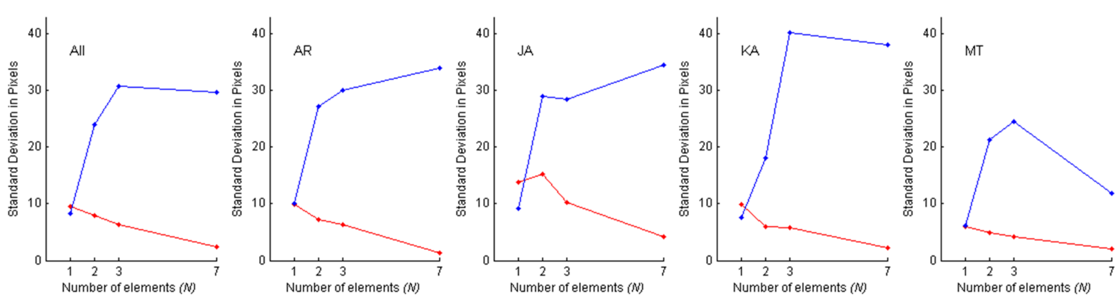

Figure 4. Values of sigma in mean size discrimination task for each number of test-set elements are marked with red dots and the values of sigma in summary size discrimination task for each number of test-set elements are given with blue dots.

Discussion

Kahneman (2011) introduced an idea about two decision making systems, one working fast and other slow. On the basis of this dual system theory, Kahneman (2013, p. 92-93) proposed that our visual system allows us to estimate average length or size of similar elements with considerable accuracy by engaging System 1. On the other hand, if estimating total length or summary size of similar objects, System 1 is extremely inefficient. In Kahneman's example, for finding the total length of a set of lines, we need to estimate the average length of the elements and then multiply it by the number of elements in the set. Multiplication, however, can be done only by cognitively more elaborate, but time consuming system, System 2. (Kahneman 2011, p.92-93).

For testing this intriguing proposal, we presented observers sets with four different sizes and numbers of elements where only the size of reference element and the instruction were changed for different experimental tasks. We expected the results to support Kahneman's hypothesis, if in similar conditions, in case of the same test stimuli, human observers are not able to estimate total size of elements as accurately as they can estimate average size of a set of similar objects. Obtained data demonstrated that accuracy in summary size estimation was several times lower than it was for the mean size estimation. Therefore, the results seem to support Kahneman's claim, that the visual system is relatively inefficient in estimating summary size of a collection of identical geometric objects. However, our data do not provide any evidence to Kahneman's other proposal that for finding the sum size it is necessary to use multiplication which is privilege of the deliberate and slow System 2.

As mentioned above, Kahneman's assumption that averaging process must be performed before summing process seems contradictory. Following mathematical definition of arithmetic mean, estimating mean size of similar elements would mean computing total size of the elements by adding them up to each other one-by-one, and then dividing these by the number of addends. The size of the resulting "image" is compared to the reference element. Therefore, if summation is presupposed in mean size estimation, it raises a question, why human observers do well in mean size discrimination tasks but not in the summary size discrimination. Moreover, it contradicts the assumption, that total size estimation is inhibited by multiplication, since dividing as a multiplicative process is presupposed as one of the underlying operations of computing average size.

That leads to a question if it is possible to calculate average size without using division. One way of avoiding division in estimating average size in mean size discrimination tasks is using recursive comparing. Instead of adding up sizes of all the elements and then dividing, differences between reference and test elements are estimated one-by-one and then recursively summed up to an accumulating value. If the sum has a negative value, decision – "is smaller" – is made, if the sum has a positive value decision –"is larger" – is made. That kind of model would allow avoiding the requirement to add division operation explicitly to the mean estimation model. It also does not matter if we accumulate differences of all N test elements from the reference or only a subset K of all N elements. However, by using the recursive formula for completing the mean size estimation task, we still have to use summation operation. So, if we are so good at finding the average size through sum, why are we not using the same operation for finding the total size of elements?

Before making any conclusions about the operations visual system applies while estimating summary size, we need to consider differences between the properties estimated in summary and mean size estimation experiments. As mentioned above, by Weber's Law, the greater is the magnitude of a certain property; the greater has to be the difference between two objects for just noticeable difference (JND). Therefore, it is very likely that the accuracy in summary size discrimination is in comparison to mean size discrimination reduced by the size difference of the properties that have to be estimated. If we are estimating mean size of a set of circles, we are estimating a property of a set that is N times smaller than the property that is used while estimating total size of the set of elements. Therefore, it is reasonable to believe that the JND in case of summary size discrimination is at least $k \times N$ times higher than JND in case of mean size discrimination where k refers to the Weber's coefficient. However, accuracy in mean size estimation is expected to be increased by $\sqrt{N}$, since statistically, increase in the number of elements leads to increase in estimation accuracy. (Allik et al. 2013)

It is rather obvious that one does not need multiplication to compute the sum size, if it is sufficient to sum sizes of all elements together. The Torgerson's conjecture also claims that human visual system is probably using additive operations in visual tasks requiring estimating ratios between visual attributes. Kahneman was evidently mistaken assuming that for estimating summary size it is necessary to multiply which can be done only involving deliberate and slow System 2. Since there is no need for multiplication there is also no need in deliberate and slow System 2. However, as it can be seen from Figure 4 the accuracy in summary size discrimination was significantly worse (up to ten times) than the mean size discrimination.

It was proposed in the introductory part that the main difference in summary and mean size estimation is the presence or absence of the direct comparison between test elements and the reference. In mean size estimation task each test element, independent of all other test elements, can be directly compared with the reference and their perceived difference in size can be measured and recorded. The direct comparison, however, is not possible in the sum size task: the reference presented on the screen is not compatible with any of the visible test elements by its size but only with an imaginary element absent from the screen. One way how to estimate sum of all test circles without assuming multiplication is to place all perceived circles side-by-side on an imaginary line (see Figure 5) and thereafter estimate the length of the line. It is important to notice that the perceived size does not necessarily coincide with the actual physical size. The perceived size could be smaller or larger than the physical size. Since all the elements are measured with an unavoidable error, the estimated sum of all the elements deviate from the real sum by the sum of dispersions of all the single elements. Thus, error in sum size estimation (σ) has to increase with the number of elements N. Therefore variance in sum size estimation as well as mean size estimation might be caused by measurement error of single elements. Since mental transposition with which test elements can be placed on an imaginary line can add an additional imprecision in the estimation of the element's size, it is expected that the internal noise (ς) with which each element can be measured is considerably higher in the sum size than in the mean size task. Indeed, if the mechanism of the sum size discrimination is in principle as it is depicted in Figure 5 then it is easy to determine internal noise ς associated with each individual element from the slope of psychometric function ($1/\sigma$). Since variance of the sum of N random variables is the sum of variances of these N random variables, the mean internal noise associated with each test elements is $\varsigma = \sigma/\sqrt{N}$. For example, if for two elements (N=2) the standard deviation of the best fitted psychometric function $\sigma = 27.28$ then $\varsigma = 19.29$ pixels. Provided that $\sigma = 34.08$ and $N = 7$ then $\varsigma = 12.88$ pixels. Thus, even if standard deviation of the psychometric function σ slightly increases with the number of elements N for sum size task (Figure 4, blue lines), in reality it means that internal noise ς with which diameter of 7 circles can be determined is smaller than when there are only 2 or 3 test circles to join together.

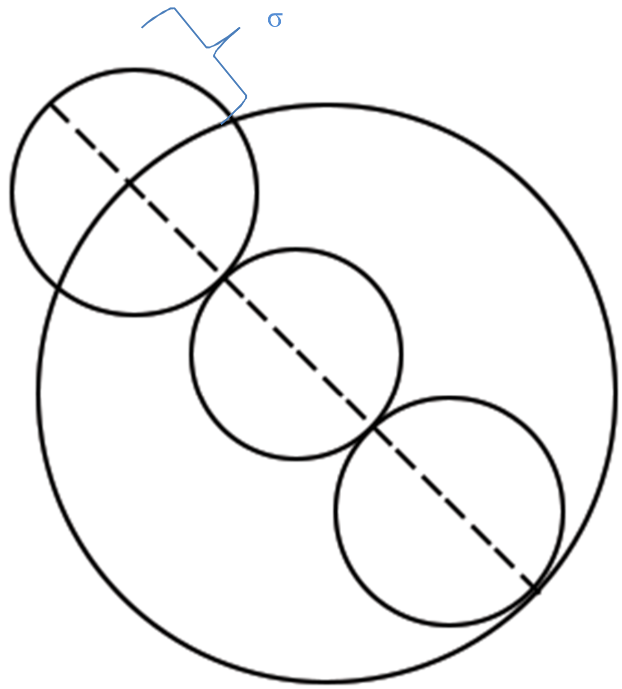

Figure 5. Sum size discrimination. For estimating the sum size, all the elements are placed to an imaginary line. The length of the line (broken line in the figure) designates the sum of all the elements that is in sum size discrimination task compared to the reference.

Once more, if sum size and mean size are both estimated through the same summation operation and the only source of error in both cases is the measurement error of a single object, why is the accuracy in sum size estimation times lower than it is in mean size estimation? One possible explanation is that the specificity of the tasks favors mean size estimation. If we are required to estimate mean size of a set of elements by recursive model of mean size estimation, we can take elements one-by-one and compare them to the reference element. As described above, the sum of the differences allows making a decision about the relationship of the mean size of the test-set elements and reference element. On the other hand, if we are required to estimate the sum size of elements we cannot compare any of the test-set elements directly to the reference. For sum size estimation, in addition to measuring all the elements one-by-one, we have to set them to an imaginary line and then compare to the reference. The additional requirement of mentally "relocating" the elements is likely to cause increase in measurement error of the elements. Moreover, we saw that mean size of a set of elements can be estimated based on only a fraction of all the elements. Based on the model proposed, for effectively estimating the sum of similar elements, we would need to use all the elements present. Such need is likely to additionally increase the error in sum size estimation by increasing the measurement error of single elements.

Conclusions

Kahneman (2011) proposed that there are two different mechanisms underlying mean size estimation and summary size estimation. If mean size can be estimated by using effortless and automatic System 1, then sum size can be estimated only by attention driven System 2, since summary size estimation requires multiplication operation which System 1 is not able to perform. Thus, we expected that in case of the same stimuli but different task, observers are considerably more accurate in mean size discrimination than they are in summary size discrimination. Our results supported this proposal. We saw that the observers were many times more accurate in estimating the mean than the sum. Unlike Kahneman, we explained the difference in accuracy not by inability to use multiplication operation. Also in the scaling experiments it was proposed that the human observer is unable to say how many times larger or smaller is the difference between two estimated targets replacing this task with addition operation (Torgerson's conjecture). We assumed that the difference is caused by requirements with different difficulty on stimuli level. In the mean size estimation task test-set elements can be compared to the reference one-by-one. Valence of the sum of the differences between the sizes of each test-set element and reference circle determines the answer. On the other hand, for sum size estimation task all the elements have to be set side-by-side to each other and then the decision about the sum distance occupied by all adjoining elements has to be made. Of course, there are many other possible methods how to determine summary size of N elements but this one seems to be one of the simplest and intuitively feasible. The sum size estimation task does not allow immediate comparisons between test and reference elements and obviously require mental operations (mental transposition for example) which may decrease the precision with which the perceived size of each element can be determined. Moreover, the task specific requirement to use all the elements is likely to be source for even greater measurement error.

In the mean size estimation task we were able to replicate the results that confirmed that not all the elements are necessarily taken into account in this task. This strategy, however, is excluded in the sum size task which in principle requires taking all elements into account. We saw that the observed accuracy in the mean size task can be achieved if only a subset of all the test-set elements are taken into account for making a decision about mean size. The precision of the sum size discrimination was much worse due to perceptual operations which are needed to compensate the absence of the possibility of the direct visual comparison between test and reference elements.

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. doi: http://dx.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. doi: 10.1016/j.tics.2011.01.003

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162. doi: 10.1111/1467-9280.00327

Birnbaum, M. H., & Veit, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics, 15*, 7-15.

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

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

Garner, W.R. . (1954). A technique and a scale for loudness measurement. *Journal of the Acoustical Society of America, 26*, 73-88.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision, 9*(9). doi: 710.1167/9.9.7

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Luce, R. D. (2012). Torgerson's conjecture and Luce's magnitude production representation imply an empirically false property. *Journal of Mathematical Psychology, 56*(3), 176-178. doi: 10.1016/j.jmp.2012.02.002

Masin, S. C. (2013). On the ability to directly evaluate sensory ratios. *Attention Perception & Psychophysics, 75*(1), 194-204. doi: 10.3758/s13414-012-0382-0

Miller, A.L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology, 81*, 16-21.

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. doi: 10.3758/pp.70.5.772

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 13, 11-11. doi: doi:10.1167/11.12.13.

Spencer, J. (1961). Estimating averages. *Ergonomics, 4*, 317-328.

Spencer, J. (1963). A further study of estimating averages. *Ergonomics, 6*, 255-265.

Stevens, S. S. (1975). *Psychophysics: introduction to its perceptual, neural, and social prospects*. New York,: Wiley.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York,: Wiley.

Torgerson, W. S. (1961). Distances and ratios in psychophysical scaling. *Acta Psychologica, 19*, 201-205.

Acknowledgements

I am grateful to our research team: especially my supervisors Jüri Allik and Aire Raidvee for their advice and guidance; Mai Toom for her comments to the first draft and help with the calculations and Mait Samuel for the help with the experiments. Not only have they guided me through the writing the thesis, but also participated in the experiments as observaers. I would also like to thank my brother, Matis Averin, for his help with the schematic figures and my friend, Asja Mihhailitsenko, for reading and commenting on the paper.

Non-exclusive licence to reproduce thesis and make thesis public

I, Kristiina Averin

(*author's name*)

(date of birth: 5. Jan. 2013),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

PERCEPTION OF THE MEAN AND SUM SIZE OF GEOMETRIC FORMS

(Tõlkes: *Geomeetriliste objektide keskmise ja summaarse suuruse tajumine.*)

(title of thesis)

supervised by Prof. Jüri Allik, Dr. Aire Raidvee,

(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **24.05.2013**