REPRESENTATION OF SIMILAR OBJECTS AS A GROUP:

EFFECTS OF VARIANCE ON ENSEMBLE REPRESENTATIONS

YELDA SEMİZER

BOĞAZİÇİ UNIVERSITY

2013

REPRESENTATION OF SIMILAR OBJECTS AS A GROUP:

EFFECTS OF VARIANCE ON ENSEMBLE REPRESENTATIONS

Thesis submitted to the

Institute for Graduate Studies in the Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

by

Yelda Semizer

Boğaziçi University

2013

Representation of Similar Objects as a Group: Effects of

Variance on Ensemble Representations

The thesis of Yelda Semizer is approved by:

Assoc. Prof. Ayşecan Boduroğlu

(Thesis Advisor)

Assoc. Prof. Burak Güçlü

Assist. Prof. Esra Mungan

June 2013

Thesis Abstract

Yelda Semizer, “Representation of Similar Objects as a Group: Effects of Variance on Ensemble Representations”

Previous research has shown that the visual system forms accurate statistical summary representations, but little research to date has directly investigated how variance is represented as an ensemble statistics. In the first set of experiments (Experiments 1, 2A and 2B), we examined the effects of group variance on ensemble statistics. We presented viewers with either a single circle/line or with displays that had nine circles/lines, and asked them to estimate the mean size of these circles/lines by adjusting the size of a test circle/line. Across conditions we manipulated variance in the displays while keeping mean size constant. In Experiments 1 and 2A, participants were equally accurate in estimating the mean size of nine same-size circles/lines and a single circle/line. However, more irregular displays of Experiment 2B resulted in more error in the estimations of the size of nine same-sized lines. More critically, in higher variance displays, participants were more likely to overestimate mean size suggesting that variance biases mean size estimations. In a second set of experiments (Experiments 3 and 4), we directly tested whether variance is represented as an ensemble statistics. Participants judged whether two consecutively presented displays consisting of circles of various sizes and yet with identical means and variance. When the second display had higher as opposed to lower variance, accuracy was significantly lower suggesting that people were overestimating the variance of the initial display. The results support the notion of the automaticity of the ensemble statistics. The overestimation of variance representations might lead to overestimation bias in the mean size estimations.

Tez Özeti

Yelda Semizer, "Benzer Nesnelerin Grup Olarak Temsili: Grup İstatistiklerinde Varyansın Etkisi"

Önceki araştırmalar görsel sistemin istatistiksel özet temsiller çıkarma yetisine sahip olduğunu göstermiştir. Ancak, az sayıda çalışma varyansı bu temsiller açısından incelemiştir. İlk deney setinde (Deney 1, 2A ve 2B), varyans bilgisinin grup istatistikleri üzerine etkileri incelenmiştir. Katılımcılara tek bir daire/çizgi veya dokuz daireden/çizgiden oluşan setler gösterilmiş ve katılımcılardan bir test dairesini/çizgisini ayarlayarak bu dairelerin/çizgilerin ortalama büyüklüklerini tahmin etmeleri istenmiştir. Koşullar arası ortalama eşit tutulurken varyans değişimlenmiştir. Deney 1 ve 2A sonuçları, katılımcıların eşit büyüklükteki dokuz dairenin ortalama büyüklüğünü, tek bir dairenin büyüklüğü kadar doğru bir şekilde temsil ettiklerini göstermiştir. Ancak, daha düzensiz görsel setlerden oluşan deney 2B sonuçlarına göre, eşit büyüklükteki dokuz dairenin ortalamasını tahmin etmek hata miktarını artırmıştır. Daha da önemlisi, varyansın yüksek olduğu durumlarda, katılımcılar ortalamayı olduğundan daha yüksek temsil etmişlerdir. İkinci deney setinde (Deney 3 ve 4), varyansın doğrudan istatistiksel özet bilgisi olarak temsili incelenmiştir. Katılımcılar art arda sunulan, aynı ortalama ve varyansa sahip fakat farklı büyüklükteki dairelerden oluşan iki görsel seti kıyaslamışlardır. İkinci set ilk sete kıyasla daha yüksek varyansa sahip olduğunda, katılımcıların performansının düştüğü saptanmıştır. Bu bulgu, varyansın olduğundan daha yüksek temsil edildiğini göstermektedir. Sonuçlar özet temsillerinin otomatikliği görüşünü destekler niteliktedir. Varyansın olduğundan daha yüksek olarak temsil edilmesi, ortalama tahminindeki yüksek temsillerin de kaynağı olmaya güçlü bir adaydır.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Assoc. Prof. Ayşecan Boduroğlu. The turning point of my life was the first time I listened to her lecture seven years ago. She introduced me to the most exciting areas of psychology and gave me the strength to enrich myself with all the possibilities. I have always been impressed by her intellectual curiosity and admired her academic skills. For the past two years, during my tough times, she was there for me, supporting and caring. I feel very lucky for having a chance to work under her supervision.

I am deeply grateful to Prof. Reşit Canbeyli for his continuous support and guidance, during this thesis and also throughout my education. I have always been impressed by his academic skills, his curiosity and excitement towards new ideas.

I owe my special thanks to Assist. Prof. Esra Mungan for her keen interest in this thesis, for giving me the chance to learn by teaching and her stimulating and constructive feedback throughout my years in the psychology department.

I would like to thank Assoc. Prof. Burak Güçlü for his genuine and invaluable comments with his critical perspective. Without his help, this thesis would not be as rich as it has been.

I would also like to thank every member of the Psychology Department of Boğaziçi University. I owe my special thanks to Prof. Ali Tekcan, Prof. Bilge Ataca, Assoc. Prof. Adil Sarıbay and Assist. Prof. Elif Aysimi Duman for their continuous support and making me feel like home.

I was very lucky to have such amazing colleagues in our office. I am grateful to Burcu Kaya Kızılöz, Hatice Yılmaz, Irmak Olcaysoy, Naziye Güneş, Serdar Metin, Pınar Kurdoğlu and Melih Barsbey for their collaboration and continuous support.

I am deeply indebted to all of my friends who have provided me with a "secure base" during my thesis. Especially, I would like to thank Ali Uzun, for his personal contribution to this thesis.

I would like to thank and acknowledge TUBITAK (Scientific and Technological Research Council of Turkey) for their financial support.

I owe my deepest thanks to my family: To my dad for teaching me that productivity is what keeps a person alive, to my mom for giving me the force to move on, and especially to my beloved sister for always being there for me. I'm indebted to them for giving me the chance to follow my dreams.

And, Yalçın, I felt your support every time I remember you saying "Whatever it takes, you are given". I will always be indebted to you for flourishing the life and for your genuine smile.

CONTENTS

TABLES

FIGURES

CHAPTER 1

INTRODUCTION

In everyday world, the environment we live in is rather complex, we are constantly faced with regularities and redundancies. For a moment, think of leaving the room you are in right now and going outside to take a walk. What would you see? Trees on the side of the road, crowds of people walking or cars passing by you. When you come back, how much information do you think you would recall? Could you accurately tell on average how green the leaves of the trees were or how fast the cars passed by you? Would you make a plausible guess about the facial expressions of people you come across while taking your walk? Or, could you identify people you encountered?

Think about another situation: Let’s say you walked through a market-place to buy some fruits. As you walked, you made quick glances to the fruits in the boxes in front of the each store; finally you decided to buy oranges from one store, yet bananas from another. What led you to make your decision? How could you tell that a particular store sells better oranges by merely looking at them for such a short period of time?

The answers of the above questions lie within a particular system; that is visual working memory (VWM). VWM is the system responsible for actively holding information in mind about our current visual environment. The limited capacity of the visual system allows us to accurately maintain only a small amount of information (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997).

However, people can extract general meaning or the “gist” information of a scene very rapidly (Olivia & Schyns, 2000; Oliva & Torralba, 2006). There is the proposed ability to represent similar objects as a set or an “ensemble” on a variety of features (for a comprehensive review, see Alvarez, 2011). An “ensemble representation” refers to a representation formed through compression of multiple sources of similar information into a comprehensive summary. When presented with a set of resembling objects, it is suggested that people automatically and readily form an “ensemble representation” of the given set and they use this information to guide their future judgments (Ariely, 2001; Brady & Alvarez, 2010; Chong & Treisman, 2003, 2005a-b). For example, people are able to compute averages of mean size (Ariely, 2001; Chong & Treisman, 2003), speed of moving objects (Emmanoil & Treisman, 2008) or orientation (Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Not only low level cognitive processes but also higher level processes can utilize ensemble representations. For example, observers were accurate in estimating the average emotion of a given crowd as well as its general gender make-up (Haberman & Whitney, 2007). Similarly, people showed high performance in categorization of scenes even at quick glances (Greene & Oliva, 2009).

Although the accuracy and structure of ensemble representations over various features (specifically, average mean size) have been investigated, the mechanisms of these representations need further examination. In this thesis, we aim to investigate (a) if people are influenced by other descriptive statistics, namely variance, while making judgments about the mean size of a given set; (b) if there is a systematic bias in these judgments; (c) whether people accurately compute ensemble representations of variance.

Ensemble Representations

As proposed by Alvarez (2011, p.122), the term “ensemble representation” can be considered as an “umbrella term” that covers literature on ‘global features’ (Oliva & Torralba, 2006), ‘holistic features’ (Kimchi, 1992) or ‘statistical summaries’ (Chong & Treisman, 2003). Alvarez stated that studies on these different domains lead to a general conclusion that ensemble representations are beneficial to visual cognition. For instance, by minimizing error in representation of each item, the summary statistics enhances information gathered from a visual display. In addition, given the capacity limitations of VWM (Alvarez & Cavanagh, 2004; Cowan, 2001), ensemble representations help us acquire a general meaning of a display, even when we are not capable of maintaining all items individually. This capacity definitely is significant in terms of survival (Greene & Oliva, 2009). For instance, calculation of average speed of approaching predators is crucial. Moreover, ensemble statistics can help form more accurate item representation by reducing the errors in maintenance through adjustments by summary representations (Brady & Alvarez, 2011).

VWM Capacity and Ensemble Representations

People maintain information about their visual environment through the VWM system. However, VWM capacity is limited (Alvarez & Cavanagh, 2004; Cowan, 2001), making it hardly possible to hold all of the available information in memory. Marois and Ivanoff (2005) defined this capacity limitation as one of the three major “bottlenecks” of information processing ability from sensation to action in the human brain (for others, see Marois & Ivanoff, 2005, p. 296). Although there is a consensus that VWM capacity is limited, the nature of this limitation recently became a subject of an active debate between two opposing camps. According to one group of researchers, VWM capacity is limited to 3-4 items, each with a fixed precision (Awh, Barton, &Vogel, 2007; Cowan et al., 2005; Luck & Vogel, 1997; Zhang & Luck, 2008). On the contrary, according to another group of researchers, VWM capacity is determined by a pool of resources which can be allocated among all stored items flexibly (Bays & Husain, 2008; Boduroglu, Mueller, Ng, & Shah, under review; Wilken & Ma, 2004). There are also other hybrid models (e.g., Alvarez & Cavanagh, 2004) which argue for flexible resource sharing among a set of items, but there is an upper limit to the maximum number of items that can be stored. Thus, these current models differ in their arguments about the number of items that can be simultaneously stored in VWM as well as the precision of each stored item. However, these models all agree that VWM capacity is limited.

The ensemble representations perspective suggests that the negative consequences of VWM capacity limitations can be ameliorated by computation of summary statistics. In other words, since VWM system is not capable of registering all available information discretely at a time, it is possible that it utilizes a different mechanism which summarizes the whole scene and gives a general idea. This summary representation can be beneficial in terms of minimizing the error which can result from representing each item separately (Alvarez, 2011), leading to a more accurate but compressed representation.

Methodological Approaches in the Mean Size Estimation Studies

A large amount of research on ensemble representations focused primarily on average size estimations as summary statistics (e.g., Ariely, 2001; Brady & Alvarez, 2010; Chong & Treisman, 2003, 2005a-b). Before reviewing the relevant literature, we believe that familiarization with the methodological approaches of this field of research is necessary. Therefore, we will provide a quick tour of three common methodological frameworks used in average size perception studies.

Mean size discrimination of previously presented array to a single probe. In this first approach, the most common one, observers perceive an array of similar disks followed by a single probe and are asked to judge if the probe is larger or smaller than the mean size of the disks in the set (Ariely, 2001). Previous studies showed that when the size of the probe differed from the mean size of the set by 4% - 12%, observers were almost 75% accurate in their mean size estimations.

Mean size discrimination of simultaneously presented arrays. More recent studies utilized a different approach in which observers are presented with two arrays on each side of the screen and asked to judge which set has the larger mean size by a forced choice method (e.g., Chong & Treisman, 2003, 2005b). Findings showed that observers were approximately 75% accurate in their judgments when the difference between displays was 6%-12%, similar to the earlier methodology. Moreover, variations in the size of the individual items did not seem to influence observers' performance. However, the effects of variance on average size estimation still require research.

Mean size reproduction. One last methodology that has been used in the average size estimation studies requires people to reproduce the average size of the previous items. Observers adjust the size of a given item on the screen to the mean size of the previous set. This methodology first utilized in a pilot study to determine whether the integer of the average size judgments is the diameter or the area of the disks (Chong & Treisman, 2003). Later, other studies used this method for both mean size (Albrecht & Scholl, 2010) and single item size estimation (Bardy & Alvarez, 2010). Compared to other methods, mean size reproduction offers more precise results (Myczek & Simons, 2008). It also enables us to see the direction of any systematic bias. Therefore, in the current work, we will utilize this last methodology in the mean size estimation task.

Nature of Ensemble Representations

The accuracy as well as the structure of the ensemble representations received attention by a lot of researchers (for a review, see Alvarez, 2011). Dan Ariely (2001) was among the first to propose the idea that the visual system represents general statistical information in such context. Ariely presented participants with a set of circles with variable diameters (see Figure 1). He investigated whether the visual system represents local features of individual items (member-identification experiments) or the global summary of the set of items (mean-discrimination experiments). The member-identification experiments required observers to make judgments about the individual item membership (Was the test circle present in the previous set?). The mean-discrimination task, on the other hand, required observers to judge the mean size of the set (Was the test circle smaller or larger than the mean size of the previous set?).

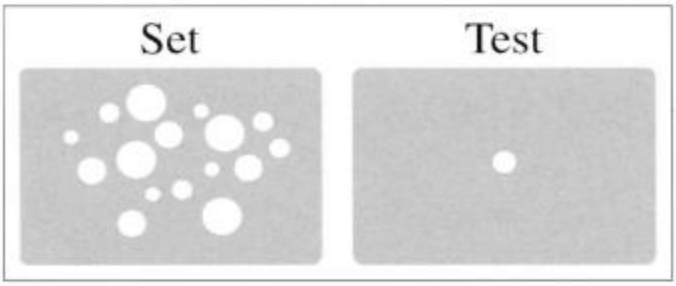


Figure 1. An example of sets used in Ariely (2001) study. Set display on the left side and test display on the right side.

Results showed that, in the mean-discrimination experiment, observers were able to judge the mean size of given spots with 75% accuracy when the size of the test circle differed from the mean size of the previous set by 4%-12%, even when presented only for 500 ms. However, in the member-identification experiment, observers were unable to make accurate judgments even if the size difference between the probe item and target item was 18%. Therefore, participants were able to judge the mean size of the sets while having little knowledge about the individual items as shown by the poor performance in the member-identification experiments. In addition, due to fast and accurate performance given such short presentation duration, averaging size perception seemed not to require much cognitive resources. However, note that both tasks depend on immediate memory. Based on these results, Ariely (2001) proposed that the visual system discards individual item information after forming the overall statistical summary.

This original study received attention of several researchers leading to the investigation of the nature of the averaging ability. Although there seems to be a consensus on its existence on several dimensions (e.g., orientation (Dakin & Watt, 1997; Parkes et al., 2001), speed (Emmanoil & Treisman, 2008) or emotion (Haberman & Whitney, 2007), there are also some inconsistencies about its characteristics. For instance, Alvarez and Oliva (2008) opposed Ariely’s (2001) idea that people discard individual item information once the overall statistical summary is formed. They instead argued that accurate information about individual items predicts the accuracy of the statistical summary representations. Therefore, whether individual item information is kept after the computation of ensemble statistics needs further research. However, this question is beyond the scope of the current work. A recent paper addressed this question and reported that accurate mean size judgments can be made without explicit registering of individual items (Corbett & Oriet, 2011). That is, observers were able to compute average size of a given set, even when they were not allowed to discriminate identities of individual items.[1]

After the work by Ariely (2001), several other studies were carried out to further establish the accuracy and the nature of ensemble representations. For one thing, Ariely's tasks depended on immediate memory. Therefore, the accuracy of mean size judgments after memory decay was still an open question. Chong and Treisman (2003) conducted a series of experiments to investigate this question. They reported that memory delays between the presentation of a set and judgment recollections only slightly affected ensemble statistics. Similarly, exposure duration of a given display had almost no effects on accuracy. In their Experiment 1, they were interested in how accurate people's judgments about average size of a set were. They simultaneously presented observers with two sets, each consisting of either 1 or 12 circles in either one or four different sizes. The task was to judge which set of circles had the larger mean size. They manipulated the distribution of items within each set to create three types of displays: heterogeneous (12 circles with varied size), homogenous (12 circles with same size) and a single circle presented alone. Two sets

[1]Here the word "compute" does not mean an effortful process. Earlier work suggested that summary representations are formed automatically (e.g., Chong & Treisman, 2005a).

presented on the screen simultaneously always belonged to the same condition. They measured the threshold of discrimination of two heterogeneous sets and compared them with homogenous and single circle displays. The threshold was defined as “the percent diameter difference between the two displays that gave %75 accuracy” (p. 396). They also varied the presentation mode (simultaneous versus successive) and the duration in the successive presentation (inter stimulus interval (ISI) as 100 ms versus 2000 ms, see Figure 2). They provided feedback for incorrect decisions by a tone.

Results showed that thresholds were found to be 6%-8% diameter difference for 75% accuracy in the simultaneous presentations, similar to Ariely’s (2001) study (see Figure 3). Also, the delay in successive presentations had only a slight effect on homogenous displays. The delay affected heterogeneous and single circle displays only when the ISI making it hard to discriminate between sets. But there were no effects of size judgment type in 100 ms or in simultaneous condition.

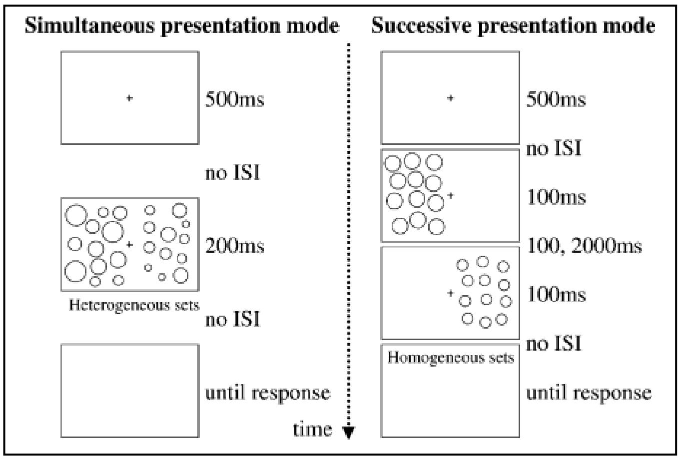


Figure 2. The timelines of Experiment 1 of Chong and Treisman’s (2003) study.

As reviewed above, contrary to what Ariely (2001) suggested, judgments about the mean size of the given set were also found to be as accurate as judgments of a single item. Another, more interesting finding of Chong and Treisman's (2003) study was concerned with the distribution of items within a set.

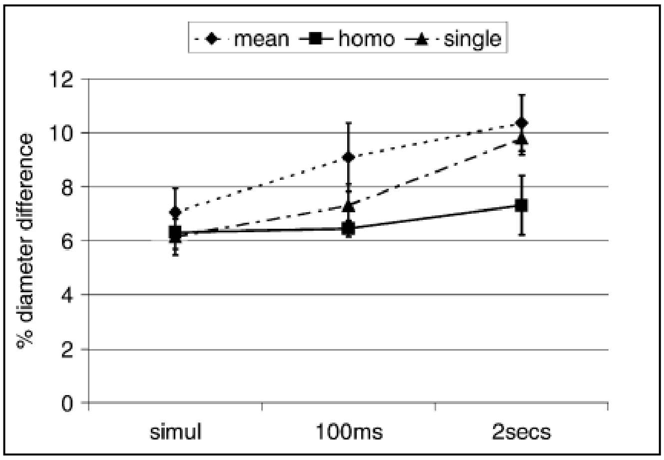


Figure 3. The results of Experiment 1 of Chong and Treisman's 2003 study.

Chong and Treisman (2003) realized that the sets used in their earlier experiments were composed of equal number of items from each four sizes (see "uniform" in Figure 4). This regularity results in accurate judgments without requiring size averaging. For example, an observer can achieve high accuracy by comparing a randomly selected item from a set to the closest size from the other set. Similarly, comparing largest items is enough to make accurate judgments. To eliminate this possibility, they conducted another experiment. They again simultaneously presented observers with two sets of circles. The task was to judge which set of disks had the larger mean size. They manipulated the distribution of items within each set to create four types of displays: uniform, two-peaks, normal and homogenous (for details, see Figure 4). The side-by-side presented displays were either identical, or were pairs drawn from these 4 types of displays.

All 10 pairs of displays were presented on the screen during the experimental session in which both expert observers who attended the earlier experiment and naïve observers participated. Results showed a small but significant difference that the thresholds for the same distribution pairs were 8% while the thresholds for different

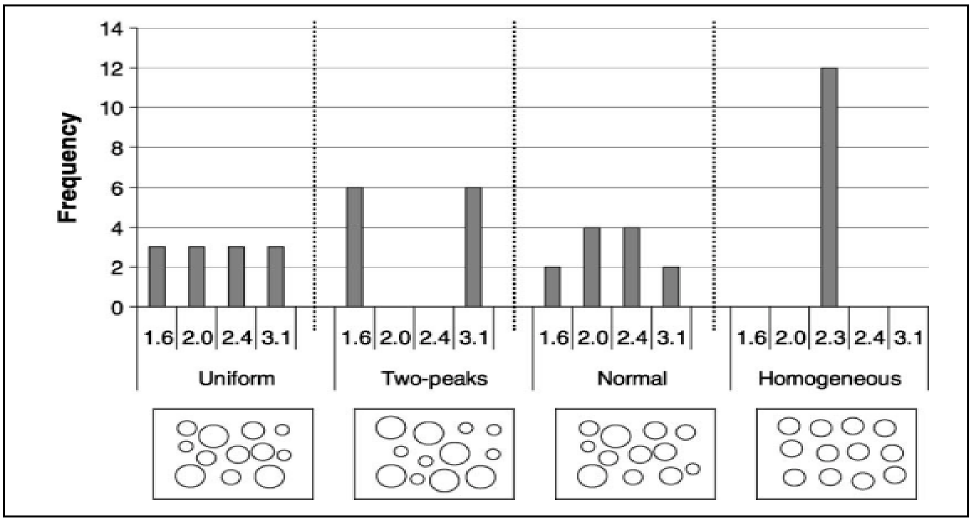


Figure 4. Four different distributions used in Experiment 3 of Chong and Treisman's (2003) study.

distribution pairs were 10% (see Figure 5). Also, two-peaks and homogenous pairs gave the highest threshold. Chong and Treisman (2003) concluded that judging sets with the same distributions provided only a slight benefit as compared to judging sets across different distributions. We think that during computation of ensemble statistics, the significance of distribution type of a set calls further research. The distribution differences used in this specific experiment is indeed very limited. The size of the circles were predetermined and limited to only four options. Also, these results do not say anything about the direction of the errors in mean judgments in

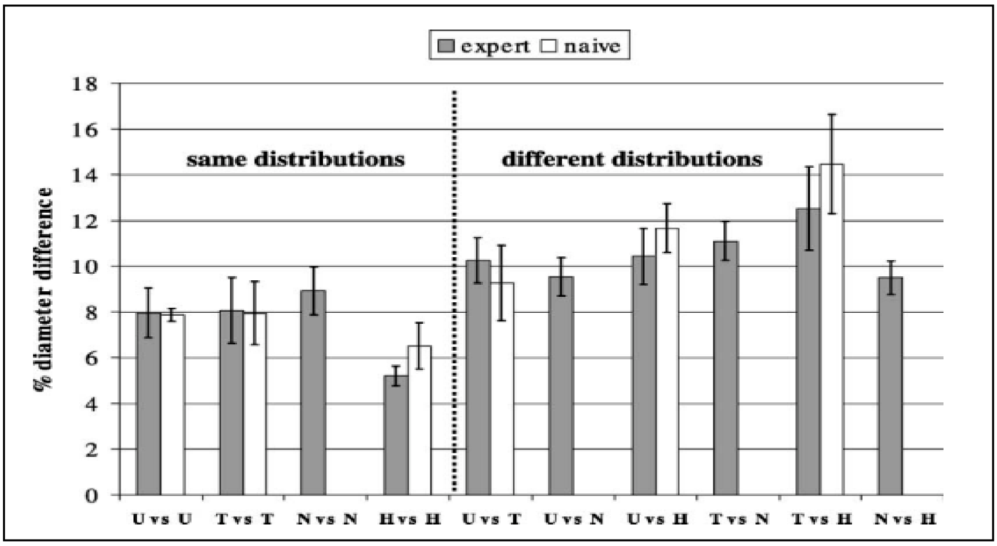


Figure 5. Results of Experiment 3 of Chong and Treisman's (2003) study.

different display conditions. Using more controlled variance differences, one could investigate whether people are influenced by the variance within a distribution while making average size estimations or they simply ignore it and end up with biased representations. One of the aims of this thesis is to investigate this question.

The Subsampling Hypothesis

According to Chong and Treisman (2005a), computation of ensemble representations was an automatic process, statistical averaging benefited from distributed rather than focused attention. However, the automaticity of the extraction of ensemble statistics was challenged by the subsampling hypothesis (Myczek & Simons, 2008). The subsampling hypothesis claims that the ability to form accurate ensemble representations does not necessitate perception of the whole sets. That is, by averaging over only a subsample of items, these statistics can also be achieved, even more accurately. Therefore, there is no need for a proposition of a new automatic mechanism. In addition, by utilizing only few items in a set, the subsampling hypothesis also argues against the requirement of distributed attention during averaging. This hypothesis proposes that focused attention mechanisms can explain the current findings. Several studies investigated the existence of subsampling during averaging (see Table 1).

Table 1. List of Studies Investigated the Subsampling Hypothesis

Study	Subsampling Hypothesis	Reasoning
Chong & Treisman (2005a)	No	Averaging is an automatic process.
Myczek & Simons (2008)	Yes	Computer simulations that mimic human perception showed that sampling of few items is enough for accurate representations.
Fockert & Marchant (2008)	Yes	Directed attention to a selected item influence mean representation.
Chong, Joo, Emmanouil and Treisman (2008)	No	Performance did not change when item was in the foveal area.
Corbett & Oriet (2011)	No	Explicit instuction to use strategies did not influence performance.
Brady & Alvarez (2011)	-	Subsampling to form two different averages seems to occur.

Myczek and Simons (2008) utilized computer simulations in order to verify the subsampling hypothesis. Use of computational models to mimic human perception was criticized in terms of lack of taking 'judgment error' into account (Ariely, 2008). Later, Simons and Myczek (2008) further specified that their critique of ensemble representations only focused on average size perception rather than other summary statistics, such as orientation or motion. They acknowledged that orientation or motion can be summarized over the whole items because they readily necessitate pooling over a number of receptors during perception. However, there is no specific mechanism dedicated to average size perception (for candidate mechanisms, see Corbett & Oriet, 2011).

Another evidence for subsampling hypothesis was proposed by Fockert and Marchant (2008). They found that when attention is directed to a selected item of a set, mean representation of the set is influenced systematically by the property of the selected item. They argued that these findings showed that extraction of statistical properties are not always automatic. We think that even if it is true that people subsample while computing ensemble summaries, it is not clear how they decide which items to include into the subset in the given brief period. The factors affecting these decisions need to be revealed; however, this is beyond the scope of the current work.

Chong, Joo, Emmanouil and Treisman (2008) attacked the subsampling hypothesis empirically. They tried to eliminate Myczek and Simons' (2008) proposed alternative explanations directly. They opposed the idea that observers use strategies while averaging. Firstly, they found that performance did not drop when different types of displays, each demanding different degree of focused attention, were presented in a mixed fashion rather than in blocks. They argued that if there was a systematic strategy-use, then performance would drop in the mixed design. Secondly, presentation of a subset of items rather than the whole display dropped performance. Also, there was no difference if these items were chosen from the foveal area or randomly. Lastly, performance was not affected by which display had the largest items. Although, Chong et al.'s first experiment yielded clear results, findings of the other experiments need further examination. Simons and Myczek (2008) proposed that if people are not using the strategies suggested in these experiments, they might be using other strategies.

Recently, more evidence against the subsampling strategy was proposed by Corbett and Oriet (2011). They explicitly gave strategies to participants to utilize and did not find any benefits. But again, Simons and Myczek's (2008) above criticism still holds. However, even if there could be a benefit of using strategies or subsampling, it does not state that people only compute averages over subsamples. Therefore, the subsampling hypothesis still needs to be further tested. The subsampling was investigated by a different perspective by Brady and Alvarez (2011). They examined whether retrieval of an individual item of a set from memory is biased towards the mean size of the items in the same color when color is relevant to the task. In their experiment, they created 15 pairs of displays composed of three red, three green and three blue circles (see Figure 6). The largest circles were one color while the smallest circles were another color. Then, they switched the colors of the largest and smallest circles and created 15 new displays. In all these displays the mean sizes were identical. Using these 15 matched displays, they were able to assess the bias in the size judgments.

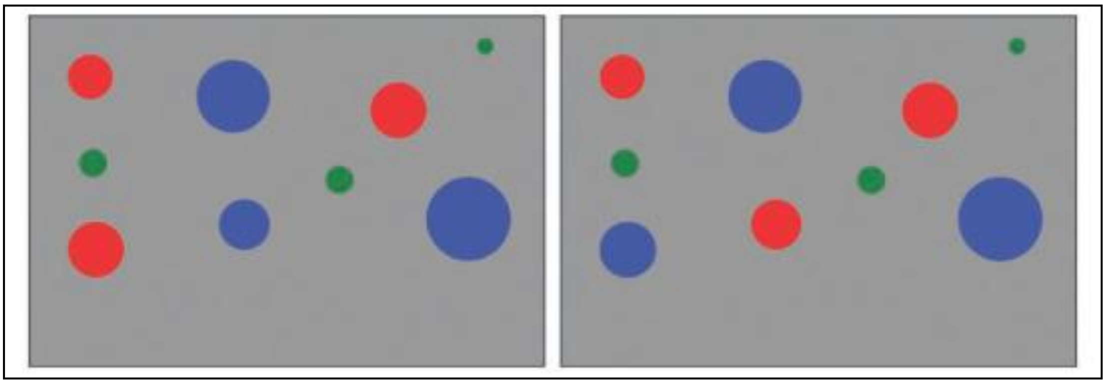


Figure 6. An example of paired displays in Brady and Alvarez's (2011) study.

Participants were presented with the same 30 displays and were told to ignore the green circles (middle-sized ones). After the display disappeared, their task was to resize a randomly sized black circle to the circle presented at the same location in the set. The bias was calculated by comparing the size estimation differences between matched displays (see Figure 7). Results showed that retrieval of an item from memory was biased towards the mean size of the items presented in the same color. Hence, ensemble representations were created automatically and manipulated individual item representations, revealed a bias towards the mean size of the items from which individual item drawn. The methodology to calculate the bias in mean size estimations in the current work will be similar to Brady and Alvarez's approach.

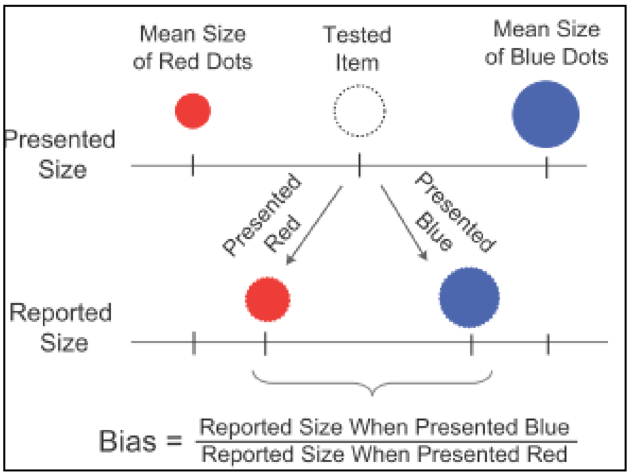


Figure 7. Bias calculation in Brady and Alvarez's (2011) study.

The Present Study

Previous research revealed that observers are capable of computing summary statistics while presented with a set of similar items (Ariely, 2001; Brady & Alvarez, 2010; Chong & Treisman, 2003, 2005a-b). Although there seems to be different views about its automaticity and existence of sampling strategies during computations, the structure and accuracy of these summary representations are established by both direct (Ariely, 2001) and indirect evidence (Van Opstal et al., 2011). However, there is still a need to reveal potential factors that can systematically bias these representations and affect outcome judgments. Similarly, other than average size estimations, the most common statistical summary measure, new descriptive parameters necessitate investigation.

Reflecting on the previous research, one of the aims of this thesis is to investigate the effects of variance on the average size estimation. We speculate that if people are affected by the variance in a given display, they will report biased average size estimations. However, if they are capable of perceiving the variance differences and correcting for them while making mean size estimations, their judgments will be similar under different variance conditions without any bias. We will conduct further analysis if there seems to be any bias to reveal its mechanisms.

Another aim of this thesis is to examine whether people accurately estimate variance of a given set of items. If people are affected by different variances in a set of items while making average size judgments, then we would speculate for a tendency to compute variance. However, in line with the earlier work (Lovie & Lovie, 1976), we would expect lower accuracy in variance judgments compared to mean judgments.

CHAPTER 2

EXPERIMENT 1

Although there were some studies on how item distributions affected mean size judgments across sets (e.g., Chong & Treisman, 2003, 2005a-b), there is little research on how variance in a given display biases mean judgments. The limited research on this particular topic utilized numbers as stimuli (e.g., Kareev et al., 2002). We believe that further examination of this aspect of ensemble statistics with visual shapes can tell us about whether people represent variance information while calculating these statistics. In Experiment 1, we tested the accuracy of mean size judgments of a given set. We also looked at the effects of working memory capacity, sensory processing sensitivity and object-spatial imagery ability on representation of ensemble information. Then, we further examined whether there was a systematic bias in these judgments due to the variance manipulation in viewed sets. We predicted that observers would generally make accurate mean size judgments. More critically though, we expected their judgments to be systematically biased as a function of variance of the display, with mean size judgments being less accurate in high variance conditions.

Method

Participants

Thirty six students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent (see Appendix A). All received course credit for their participation.

Apparatus and Stimuli

All experiments, stimuli, timing operations and data collection were controlled by a PC running E-prime 1.2 software (Psychology Research Tools, Inc). The stimuli were presented on a gray background on a 17-in CRT computer monitor with a viewing distance of approximately 57 cm. The resolution was set to 640 x 480. Hence, 1 pixel was approximately .05° of visual angle.

There were a total of 120 displays. Eighty of these displays consisted of 9 white circles on a gray background where we manipulated the mean and variance to create four different conditions (see Table 2): smaller mean – low variance, smaller mean – high variance, larger mean – low variance, larger mean – high variance. In addition to these, there were 20 displays with 9 same-sized circles (homogenous displays) and 20 displays with a single circle (single-circle displays). For both homogenous and single size displays, the mean size of the circles was identical to those in the low and high variance displays.

The diameter of each individual circle in the high variance conditions (for both small and large mean) was randomly selected from a normal distribution spanning the interval of 15 to 95 pixels (0.75° – 4.75°). For the small mean – low variance displays, the range of the distributions was 15-55 (0.75° – 2.75°) and for the large mean – low variance, it was 55-95 (2.75° – 4.75°). None of the individual circles had the size of the mean. Also, there was no repetition size within a set. Only items that meet our specified mean and variance criterions were selected.

The position of circles within a set was randomly determined in an invisible 5 x 4 grid; with a jitter of +/- 10 pixels to minimize regularities. Hence, the center of the screen remained empty. No more than two circles were placed adjacently. There were up to three circles either in a row or column. In order to avoid occlusion of stimuli with the edges of the screen, the matrix was positioned with a restriction of 50 pixels (2.5°) away from the each four side of the screen.

Table 2. The Diameter Ranges of Different Conditions in Pixels

	Condition	Diameter Range (pixel)[2]			
Mean Size		Mean Size	Variance (*SD*)	Individual Circle Size[3]	No of Trials
Small Mean	Single	36-44	-	36-44	10
	Homogenous	36-44	0	36-44	10
	Low Variance	36-44	5-10	15-55	20
	High Variance	36-44	20-25	15-95	20
Large Mean	Single	66-74	-	36-44	10
	Homogenous	66-74	0	36-44	10
	Low Variance	66-74	5-10	55-95	20
	High Variance	66-74	20-25	15-95	20

[2]One pixel was approximately .05° of visual angle.

[3]Because of lower probability of generating sets with low variance in high range, we used different ranges while generating individual circles within a set. Only items that our specified mean and variance criterions were selected.

Mean Size Estimation Task

Each trial started with a fixation cross presented at the center of the screen for 500 ms (see Figure 8). Following fixation, a set of 9 circles (or single-circle display) appeared for 1000 ms followed by a 1000 ms blank, after which a single randomly sized circle appeared in blue at the center of the screen. In each trial the size of the blue response circle was determined in the following way by E-prime software (Psychology Research Tools, Inc): A number between 13 and 17 was randomly determined and subtracted from the mean size of the display. Then, another number from the same range was chosen randomly and added to the mean radius of the circles. Finally, a random number between these newly calculated radii was chosen as a radius of the response circle.

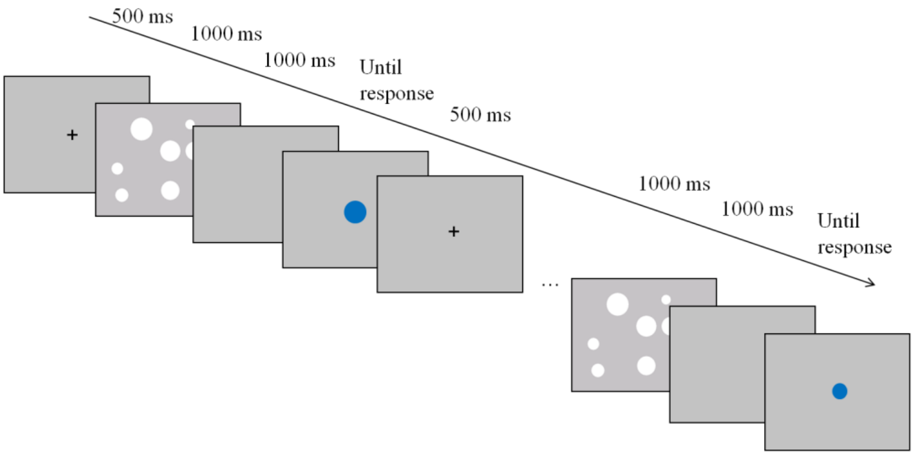


Figure 8. The design of mean size estimation task in Experiment 1.

Participants were required to resize the response circle to the mean size of the previous set via pressing ‘k’ and ‘d’ keys to either increase or decrease the circle size, respectively. Each key press increased or decreased the radius of the circle by 1 pixel (0.05°). Then, they pressed another key to record their response. Conditions were presented in a mixed fashion. Two different trial lists were randomly created with a constraint that sets with either the same mean size or variance did not appear consecutively. There were 12 practice trials at the beginning of the task. Feedback was provided only during practice trials by presentation of a circle with the correct mean size at the top of the reproduced circle.

Working Memory

An automated version of the operation span task (Ospan; Unsworth, Heitz, Schrock, & Engle, 2005) was used to measure working memory capacity. The Ospan task is proposed to represent both storage and processing capacity of working memory (Engle, Kane, & Tuholski, 1999). In Ospan, participants were asked to solve simple mathematical questions while they remembered a set of letters for later recall. There were three sections of practice trials. In the first section, participants practiced with either 2 or 3 letters, one at a time for 1000 ms and then they were asked to recall the letters in the exact order. They responded by clicking the letters they saw among the 12 letters (F, H, J, K, L, M, N, P, R, S, T, Y) presented to them, followed by clicking an “Exit” button to start the next trial. There were 4 trials in this section and feedback about how many letters were correctly recalled was provided after each trial.

In the second section, participants practiced with solving math operations. They were given a simple math operation (e.g., (4*3) + 4 = ?) at the center of the screen. After solving it they were asked to click. Then they were given an answer (e.g. 4) and asked if it was correct. They responded by clicking one of the two buttons on a screen. They were instructed to respond as accurately and as quickly as possible. There were 15 practice trials and feedback was provided after each trial. The average response time of each participant was measured to be used as a baseline (average RT + 2.5 *SD*) in the experimental trials. This way, it was possible to account for the individual differences in response times.

In the third practice session, participants practiced with both tasks simultaneously. First, they solved a series of the math problems, each followed by a letter. Then they recalled the letters in the exact order by clicking on them from 12 letters. In order to prevent verbal rehearsal, participants were told that if their response time on the math component was slower than their average time in the math practice session, they would not be allowed to answer the question and the trial would be counted as an error. There were 3 practice trails in set size 2 and no feedback was provided this time. Only the percentage of correct math responses made until that trial was presented in red at the right-hand corner of the screen and participants were told to aim for 85% accuracy.

The experimental trials were the same as those of the third practice session. The set sizes ranged from 3 to 7, with 3 sets from each set size. Therefore, there were 15 trials (75 letters and 75 math problems) in total in the experimental section. During the whole experiment, all responses were made by clicking the left mouse button. We calculated the Ospan score which was the sum of the number sets in which all letters were recalled. The completion of this task took approximately 20-25 minutes.

Sensory Processing Sensitivity

Sensory processing sensitivity (SPS, Aron & Aron, 1997) is defined as “a temperament/personality trait characterized by sensitivity to both internal and external stimuli, including social and emotional cues” (Jagiellowicz et al., 2010, p. 38). SPS involves both high sensitivity to subtle stimuli and high arousal by external stimuli (Benham, 2006). Aron and Aron (1997) developed a measure of this trait which has a high reliability and validity. In this context of ensemble representations, we are interested in sensitivity to subtleties in the environment. We believe that this trait has a potential to explain individual differences which could possibly be observed in the ability to compute ensemble statistics.

Highly Sensitive Person Scale (HSP Scale, Aron & Aron, 1997) was used to measure sensory processing sensitivity. The HSP Scale is a 27-item self-report scale with response options ranging from 1 (not at all) to 5 (extremely) with higher scores indicating greater sensitivity. The discriminant and convergent validity (Aron and Aron, 1997) and internal consistency of the HSP Scale have been shown to be strong. Cronbach’s alphas have been found to be 0.85 or higher (e.g., Aron & Aron, 1997; Aron et al., 2005; Benham, 2006).

To our best knowledge, there are no studies on the validity and reliability of this scale in Turkey. For this thesis the scale was translated into Turkish via back-translation method (see Appendix B). Example items include ‘Are you easily overwhelmed by strong sensory input?’, ‘Are you deeply moved by the arts or music?’, ‘Do other people’s moods affect you?’ and ‘Do you startle easily?’.

Object-Spatial Imagery

Visual imagery ability was generally considered as a unitary ability in the earlier literature (e.g., Paivio, 1983). Individuals were classified as good or bad imagers based on their self-ratings on a questionnaire, such as the Vividness of Visual Imagery Questionnaire (VVIQ, Marks, 1973). A more recent group of studies showed that visual imagery ability is in fact a multi-component construct, including two distinct subsystems: object imagery and spatial imagery (Blajenkova et al., 2006; Kozhevnikov et al., 2005). Object imagery is defined as "representations of the literal appearances of individual objects in terms of their precise form, size, shape, color and brightness" spatial imagery is defined as "relatively abstract representations of the spatial relations amongst objects, parts of objects, locations of objects in space, movements of objects and object parts and other complex spatial transformations" (Blajenkova et al., 2006, p. 239). Blajenkova et al. developed a self-report imagery questionnaire which measures these two subsystems separately. In terms of ensemble representations, we think that the preference for representing objects precisely in their real form, size, etc. may be a potential contributor in the formation of ensemble representations.

Object-Spatial Imagery Questionnaire (OSIQ, Blajenkova et al., 2006; Kozhevnikov et al., 2005) was used to measure visual imagery preferences. The OSIQ is a 30-item self-report scale with response options ranging from 1 (totally disagree) to 7 (totally agree) with higher scores indicating more preference. This scale consists of two distinct subscales, object- and spatial-subscale, each composed of 15 items presented in an intermixed fashion. The construct validity of the scale and the internal and convergent and divergent validity of subscales were previously established (Blajenkova et al., 2006). The internal reliability was 0.83 for the object scale and 0.79 for the spatial scale (Cronbach's alpha). Later studies using Norwegien samples reported similar reliability scores for both object (0.89) and spatial (0.81) subscales, reflecting a cross-cultural reliability (Price, 2009). The ecological validity was also determined. For example, Blajenkova et al. (2006) reported that individuals with visual art professions scored higher than those from science or humanities on object scale. On the other hand, individuals with science professions scored higher than those from visual art or humanities on spatial scale.

There are no studies on validity and reliability of this scale in Turkey. Therefore, OSIQ was translated into Turkish via back-translation method (see Appendix C). Example items for object imagery scale include 'My mental pictures are very detailed precise representations of the real things' and for spatial imagery scale items include 'My images are more like schematic representations of things and events'.

Procedure

Participants completed a Mean Size Estimation Task followed by the Operation Span Task (Unsworth, et al., 2005). They then filled out a demographics form, the Highly Sensitive Person Scale (Aron & Aron, 1997) and the Object-Spatial Imagery Questionnaire (Blajenkova et al., 2006; Kozhevnikov et al., 2005). The order of these two scales was counterbalanced. Finally, they were debriefed.

Results

In all analyses, our primary measure was the estimated radius of the circles in terms of pixels. The alpha level in all analyses was set to .05. Before analyzing data, we wanted to test whether the initial size of the response circle was equally likely to be larger or smaller than the mean size of the set in each condition. The size of the response circle was randomly determined within a restricted range for each participant (for the details please see the method section). If there was a bias towards one size, then this could affect our findings.

In order to control for this possibility, we summed up the number of trials where the test circle was initially larger and smaller than the mean size of the set for each trial. Then, we calculated the average for each condition (see Table 3). A series of within subjects *t*-tests was conducted. Results showed that only in the small mean-high variance condition, there were more trials where the response circle was smaller (M = 10.61, SD = 1.76) than it was larger (M = 8.69, SD = 1.74), $t(35) = -3.37$, p = .002. In all other conditions, the distributions of initial circles were roughly similar (all *t*s < 1.45, all *p*s > .15). Although, the initial size of the response circle was smaller in this condition, this works against our findings which are summarized below.

Table 3. Mean and Standard Deviations of the Number of Larger and Smaller Initial Radii of the Response Circle than Mean Size of the Set

		Initial Radius			
		Large		Small	
		M	*SD*	*M*	*SD*
Large Mean	Single	4.92	1.79	4.72	1.78
	Homogenous	4.78	1.48	4.92	1.42
	Low Variance	9.72	2.12	9.67	2.29
	High Variance	9.67	2.19	9.75	2.23
Small Mean	Single	4.97	1.72	4.72	1.60
	Homogenous	4.56	1.66	5.17	1.63
	Low Variance	10.11	2.31	9.06	2.20
	High Variance	8.69	1.74	10.61	1.76

Absolute Error Analysis

We calculated the absolute error by computing the absolute value of the difference between the estimated size and the correct size in each trial (radius in pixels). First, in order to examine whether estimating the size of a single circle is as accurate as estimating the mean size of 9 same-sized circles (homogenous condition), we compared the single and homogenous displays in each mean size condition. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was no main effect of display type, $F(1,35) = 2.55$, $MSE = .58$, $p = .12$, $\eta_p^2 = .07$. Error in the single displays ($M = 2.42$, $SD = .85$) was not different from the error in the homogenous displays ($M = 2.63$, $SD = .93$). However, there was a main effect of mean size, $F(1,35) = 42.71$, $MSE = .76$, $p < .001$, $\eta_p^2 = .55$. Participants made more error in the large mean size condition ($M =$

3.00, *SD* = .99) compared to the small mean size condition (*M* = 2.05, *SD* = .79). Moreover, there was no interaction effects, $F(1,35) = .87$, $MSE = .51$, $p = .36$, $\eta_p^2 = .02$ (see Figure 9).

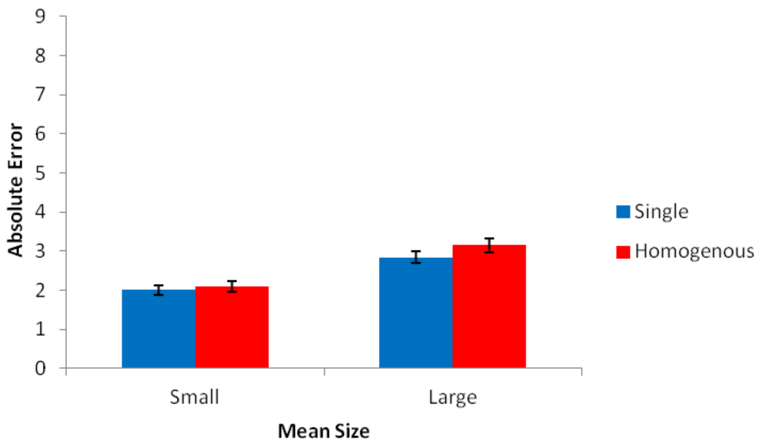


Figure 9. Absolute error in pixels single-homogenous and small-large mean size conditions in Experiment 1. Error bars indicate SEMs.[4]

Second, in order to examine whether variance affects mean size estimation, we compared low and high variance conditions in each mean size condition. A 2 (Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. As expected, there was a main effect of variance, $F(1,35) = 49.91$, $MSE = 1.23$, $p < .001$, $\eta_p^2 = .59$. Participants made more error in the high variance condition (*M* = 4.68, *SD* = 1.40) compared to the low variance condition (*M* = 3.37, *SD* = .92). Moreover, there was no main effect of mean size, $F(1,35) = 2.79$, $MSE = 1.28$, $p = .10$, $\eta_p^2 = .07$. Error in the small mean size (*M* = 3.87, *SD* = 1.05) condition was not different from the error in the large mean size condition (*M* =

[4]Error bars in all graphs indicate the Standard Error of the Mean (SEM).

4.19, *SD* = 1.27), and the interaction effect between Variance and Mean Size variables did not reach significance, $F(1,35) = .98$, $MSE = .69$, $p = .33$, $\eta_p^2 = .03$ (see Figure 10).

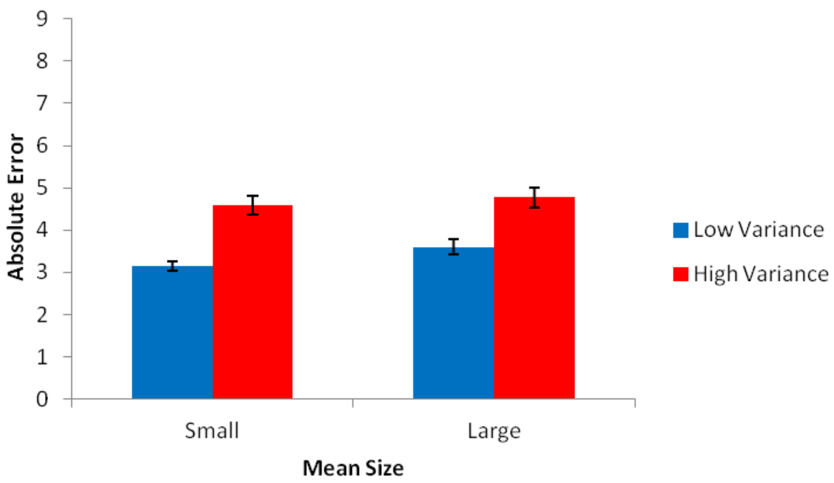


Figure 10. Absolute error in pixels in low-high variance and small-large mean size conditions in Experiment 1.

Estimation Bias Analysis

In order to reveal whether there was a systematic bias in these estimation errors, we calculated an estimation bias score by computing the ratio of estimated size over correct size. Estimation bias score greater than 1 would indicate an overestimation tendency while estimation bias score less than 1 would indicate an underestimation tendency. When compared to 1 by a series of single t-tests, only in the high variance condition, there was a significant effect, suggesting an overestimation of mean size in the high variance condition, $t(35) = 3.62$, $p = .001$ (see Figure 11 & 12).

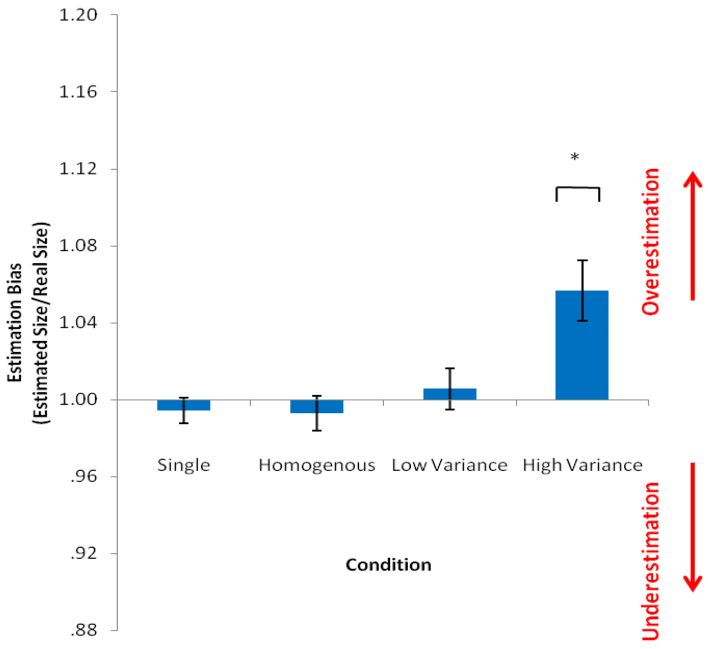


Figure 11. Estimation bias in each condition in Experiment 1.

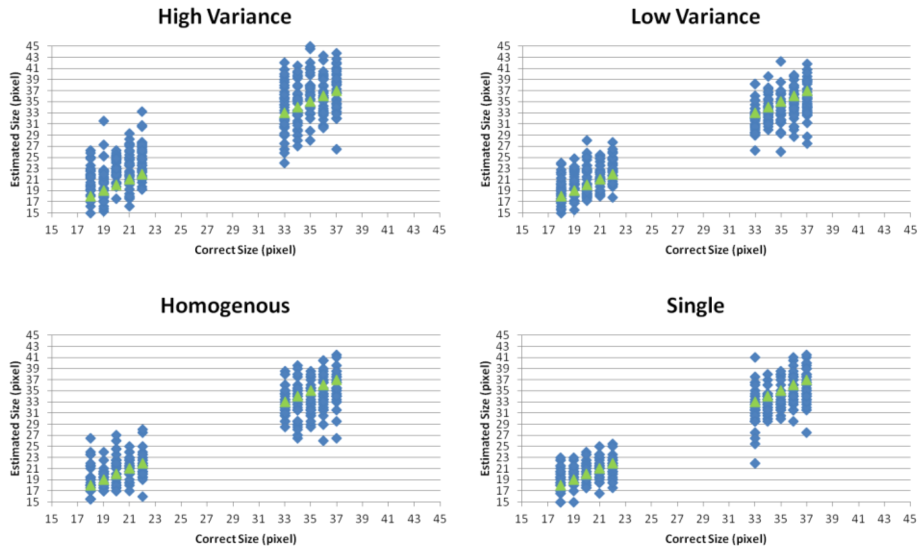


Figure 12. Estimated size (radius in pixels) as a function of correct size in each condition in Experiment 1 (Triangles indicate the correct size).

We speculated that the overestimation in high variance condition might result from the nature of the sets used in this condition. If the sets include higher number of larger circles, then this might bias the estimations. In order to investigate this possibility, for each set, we counted the number of individual circles which had larger or smaller size than the mean size of the set. The range of larger and smaller circles changed from 2 to 7 in each set. In total, there were 18 out of 40 sets which included higher number of larger circles than the mean size of the set in the high variance condition. Similarly, there were 19 out of 80 sets which included higher number of larger circles in the low variance condition. Therefore, two variance conditions were similar in terms of the number of sets with larger circles than their specified mean set size.

Real Error Analysis

We also calculated the real error in each condition by subtracting participants' response from correct size in each trial. Real error greater than 0 would indicate an overestimation tendency while real error less than 0 would indicate an underestimation tendency. When compared to 0 by a series of single t-tests, only in the high variance condition, there was a significant effect, suggesting an overestimation of mean size in the high variance condition, $t(35) = 1.22$, $p = .005$ (see Figure 13). Compared to estimation bias analysis, the only difference was a marginal underestimation tendency in the single condition, $t(35) = -1.85$, $p = .073$. Again, in the homogenous and low variance conditions, there was no significant effect ($t(35) = -1.72$, $p = .094$; $t(35) = -.52$, $p = .61$). The results suggest that when we compressed all error participants made within a condition while ignoring the specific trials, we still see an overestimation tendency in the high variance condition. This pattern of results was consistent with findings from the absolute error analyses (see Figure 11).

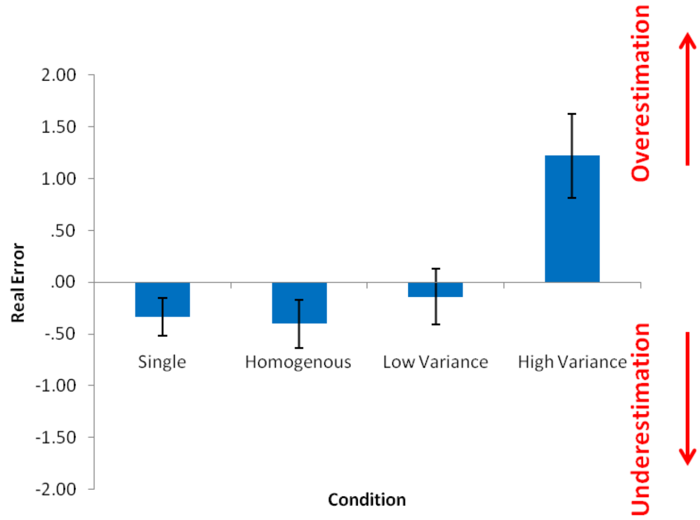


Figure 13. Real error in each condition in Experiment 1.

Coefficient of Variation

Coefficient of Variation (CV) was computed by dividing the standard deviation by the mean of the participants' responses in each condition. A series of within subjects *t*-test was conducted. First, we compared CV in the homogenous and single conditions. In the large mean size condition, there were no significant differences in terms of CV in the homogenous condition (*M* = .11, *SD* = .04) and single condition (*M* = .10, *SD* = .03), *t*(35) = .95, *p* = .35, Cohen's *d* = .28, effect size *r* = .14. Similar pattern of results was found for the small mean size (*M*s = .13, *SD*s = .04 and .05, respectively), *t*(35) = .75, *p* = .46, Cohen's *d* = .00, effect size *r* = .00. However, when we compared the high and low variance conditions, we found that in the large mean size CV was higher in the high variance condition (*M* = .14, *SD* = .04) compared to low variance condition (*M* = .12, *SD* = .04), *t*(35) = 3.12, *p* = .004, Cohen's *d* = -.50, effect size *r* = -.24. Exactly the same results emerged in the small mean size, (*M*s = .23 and .18, *SD*s = .07 and .04, respectively), $t(35) = 4.58$, $p < .001$, Cohen's $d = .88$, effect size $r = .40$. Therefore, the variations in participants' responses while estimating of the size of a single circle and the mean size of 9 same-sized circles were similar. However, their responses showed more variability in high variance conditions compared to the low variance condition for both mean size.

Individual Differences Analysis

Visual Working Memory Capacity

Participants' Ospan score changed between 18 and 75 with an average of 53. In order to examine whether participants' visual working memory capacity was related to their performance, we conducted a series of correlational analysis between Ospan score and absolute error in each condition. In contrast to our expectations, WM capacity does not seem to be related to the performance in any of the conditions (all *r*s < 2.90, see Figure 14).

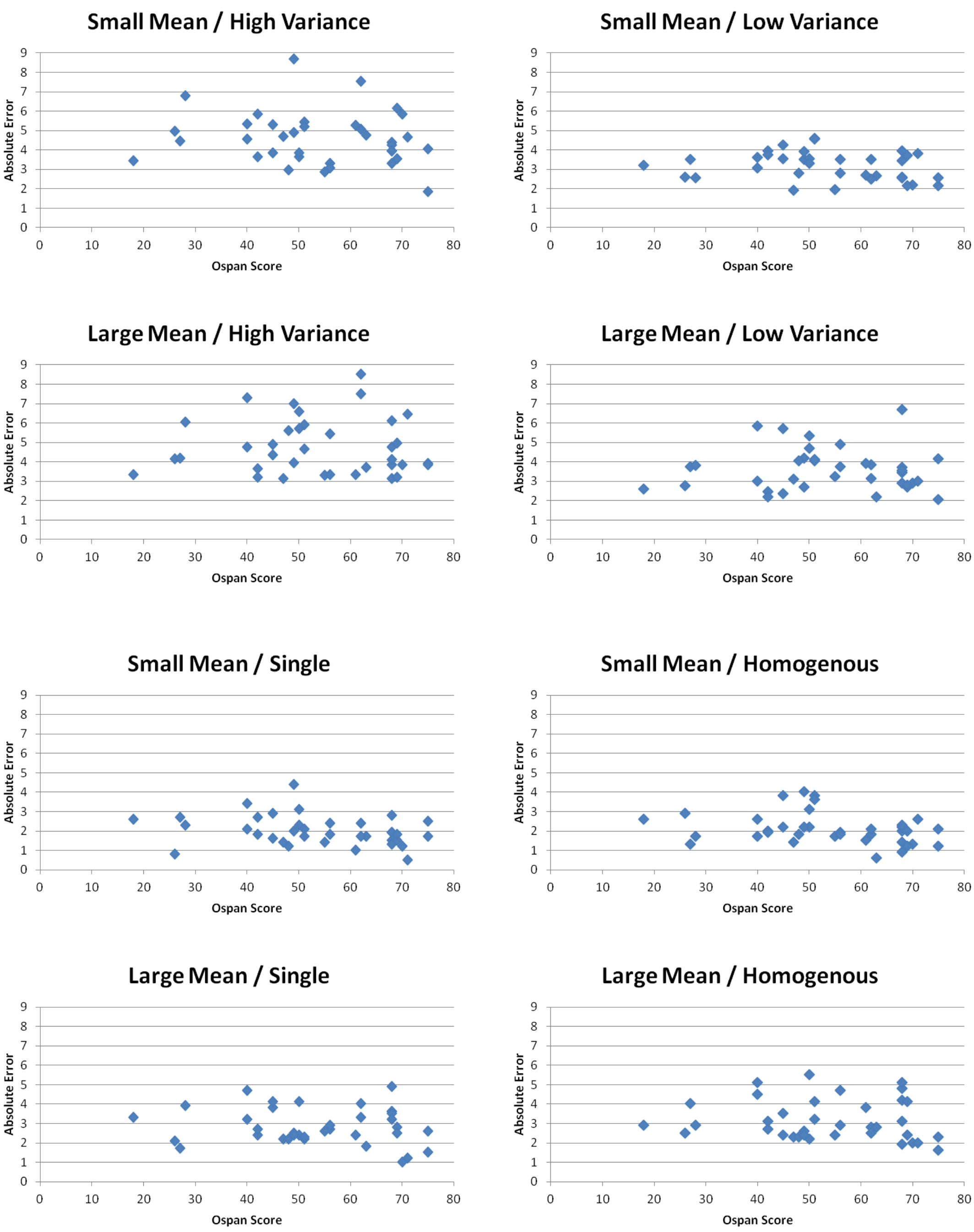


Figure 14. Absolute error as a function of Ospan score in each condition in Experiment 1.

<u>Sensory Processing Sensitivity and Mental Imagery Preference</u>

In order to examine whether individual differences play a role on the participants' performance, we conducted a series of correlational analysis (see Table 4 for descriptive statistics). HSP scale score was calculated by averaging over the participants' responses to all 27 items. Results revealed that participants' estimation bias and their score on HSP were positively correlated, suggesting that as the sensory sensitivity increased, participants reported more biased responses. OSIQ object and spatial scores were calculated by averaging over items corresponding to each subscale (15 items for each). However, scores on neither OSIQ object scale nor OSIQ spatial scale were correlated with estimation bias (see Table 5). The results suggest that sensory processing sensitivity or object/spatial imagery did not play a role in the mean size estimation performance.

Table 4. Descriptive Statistics for Scale Measures Used in Experiment 1

	Mean	SD	Minimum	Maximum
HSP Scale	3.31	0.42	2.44	4.22
OSIQ Object Scale	3.33	0.51	2.29	4.20
OSIQ Spatial Scale	3.03	0.50	1.73	4.33

Table 5. The Pearson Product-moment Correlations among HSP Scale, OSIQ Object Scale, OSIQ Spatial Scale and Estimation Bias in Each Condition

	HSP Scale	OSIQ Object Scale	OSIQ Spatial Scale	High Variance	Low Variance	Homogenous	Single
HSP Scale	-	.147	-.078	.331*	.612**	.482**	.459**
OSIQ Object Scale		-	.372*	-.207	-.150	.017	-.102
OSIQ Spatial Scale			-	-.038	-.006	.029	-.039
High Variance				-	.725**	.654**	.330*
Low Variance					-	.733**	.534**
Homogenous						-	.369*
Single							-

*p < .05, **p < .01.

RT Analysis of the First Key Press

We were interested in whether there were RT differences between the conditions. Since the size of the response circle was randomly generated in each trial, the number of key presses required changed randomly. Therefore, the total RT during each response was not meaningful. But we also logged the data for each key press in each trial. Although we did not have any specific predictions about the response times, our design allowed us to compare the time passed before first response was made in each condition. We speculated that this might imply the time participants take to adjust the size in their mind before making any response. In order to examine the RT effects, we therefore calculated the average RT for the first key press in each condition. For each participant, trials in which the RT was +/- 2 *SD* of the mean were excluded. Also, there were two outlier participants who had RTs 2 or 3 *SD* above the mean.[5] After excluding these outliers, we conducted the analysis over 34 participants. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a main effect of display type, $F(1,33) = 5.86$, $MSE = 11729.86$, $p = .021$, $\eta_p^2 = .15$. RT in the single displays ($M = 639.19$, $SD = 187.98$) was slower than the RT in the homogenous displays ($M = 594.22$, $SD = 150.61$). Also, there was a main effect of mean size, $F(1,33) = 13.95$, $MSE = 9155.10$, $p = .001$, $\eta_p^2 = .30$. Participants were slower in the large mean size condition ($M = 647.35$, $SD = 184.30$) compared to the small mean size condition ($M = 586.06$, $SD = 154.30$). Moreover, there was no interaction effects, $F(1,33) = .97$, $MSE = 8444.17$, $p = .33$, $\eta_p^2 = .03$ (see Figure 15).

[5]We utilized the same procedure for the RT data in Experiment 2A and B.

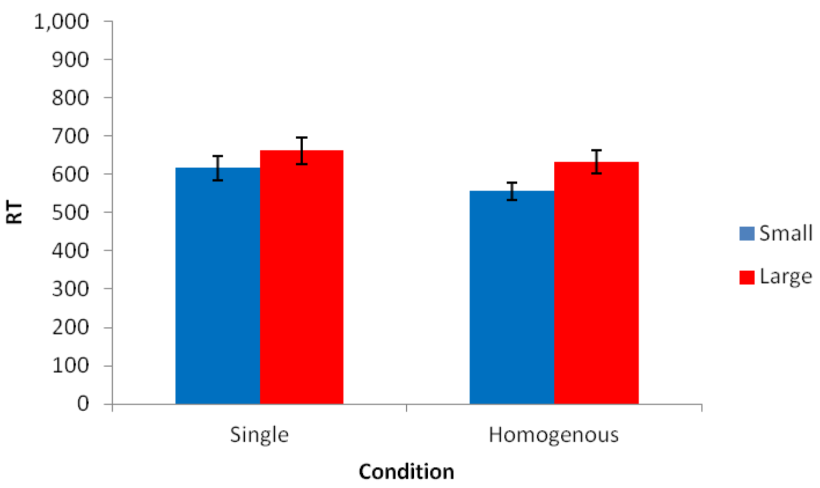


Figure 15. Average RT of the first key press in single-homogenous and small-large mean size conditions in Experiment 1.

Second, in order to examine the RT differences in different variance conditions, we compared low and high variance conditions in each mean size condition. A 2 (Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a main effect of variance, $F(1,33) = 6.70$, $MSE = 11695.47$, $p = .014$, $\eta_p^2 = .17$. Participants were slower in the high variance condition ($M = 689.04$, $SD = 230.41$) compared to the low variance condition ($M = 641.04$, $SD = 187.01$). Moreover, there was a marginal main effect of mean size, $F(1,33) = 3.45$, $MSE = 7110.55$, $p = .072$, $\eta_p^2 = .01$. RT in the small mean size condition ($M = 678.60$, $SD = 210.86$) was marginally slower than the RT in the large mean size condition ($M = 651.60$, $SD = 206.56$), and the interaction effect between Variance and Mean Size did not reach significance, $F(1,33) = 2.11$, $MSE = 6052.12$, $p = .16$, $\eta_p^2 = .06$ (see Figure 16).

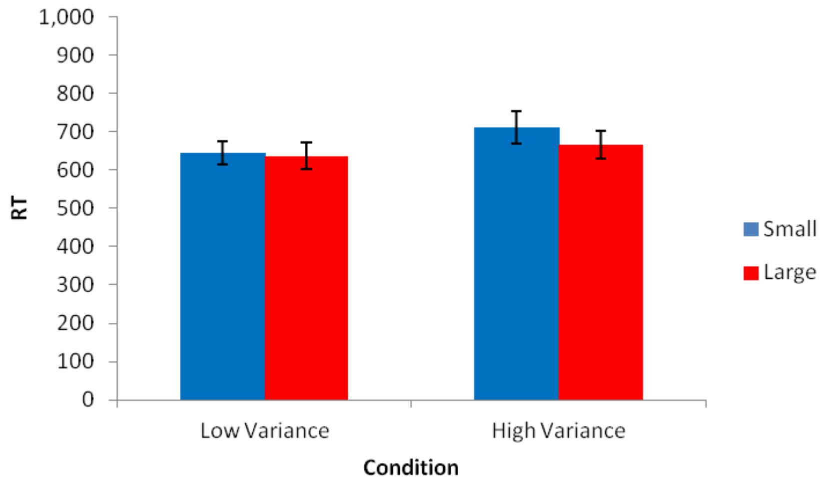


Figure 16. Average RT of the first key press in low-high variance and small-large mean size conditions in Experiment 1.

Experiment 2A

The unit of averaging in mean size estimation was investigated in a pilot study by Chong and Treisman (2003). The findings suggested that people somewhat shift their integers between the diameter and the area of the disks while averaging. In Experiment 2A, we readdressed this question in a more controlled way. We used the stimuli and procedure in the Experiment 1 with one exception: We presented participants with the diameters of the circles rather than the whole circles and asked them to average over these lines (actually diameters). Based on the previous literature, we speculated that participants' performance will be similar to that in Experiment 1.

Method

Participants

Twenty students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation.

Apparatus and Stimuli

We used the same stimuli and apparatus in the Experiment 1 with one exception (see Figure 17): We presented participants with the diameters of the circles rather than the whole circles and ask them to average over these lines (actually diameters).

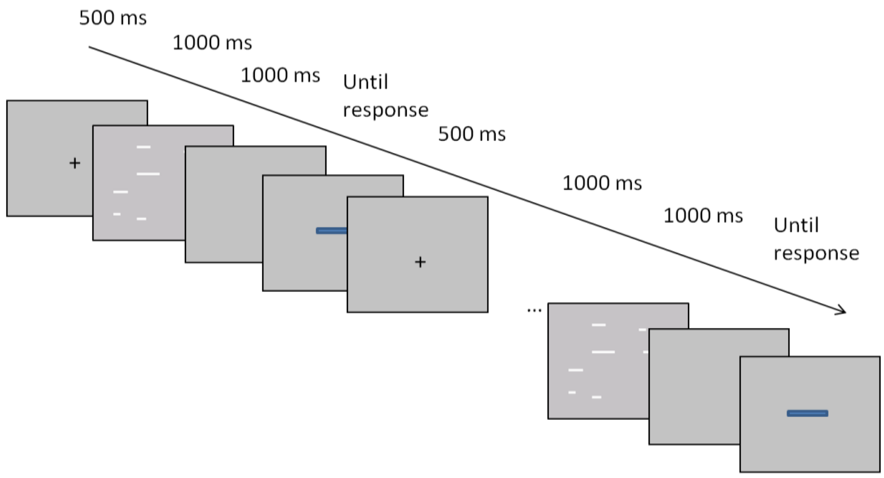


Figure 17. The design of Experiment 2A.

Results

Before any further analysis, a series of within subjects *t*-tests was conducted to test whether the size of the response circle was equally likely to be larger or smaller than the mean size of the set in each condition (see Table 6). Results showed that the distributions of the initial circles were roughly similar in all conditions (all *t*'s < 1.70 (-1.70, sign ignored), all *p*'s > .10). Therefore, the pattern of our findings was not due to the distribution of the initial circles.

Table 6. Mean and Standard Deviations the Number of Larger and Smaller Initial Radius of the Response Circle than Mean Size of the Set

		Initial Radius			
		Large		Small	
		M	*SD*	*M*	*SD*
Large Mean	Single	5.33	1.74	4.25	1.85
	Homogenous	4.83	1.49	4.92	1.47
	Low Variance	9.75	2.35	9.71	2.37
	High Variance	9.88	1.87	9.38	1.88
Small Mean	Single	4.75	1.33	4.83	1.24
	Homogenous	4.75	1.75	4.92	1.69
	Low Variance	10.08	2.47	8.96	2.12
	High Variance	8.71	2.24	10.54	2.08

Absolute Error Analysis

In order to examine whether estimating the length of a single line is as accurate as averaging over lengths of 9 same-sized lines (homogenous condition), we compared the single and homogenous displays in each mean size condition. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. Results showed a similar pattern to that of the Experiment 1. There was no main effect of display type, $F(1,19) = 2.61$, $MSE = .89$, $p = .12$, $\eta_p^2 = .12$, suggesting that error in the single displays ($M = 2.68$, $SD = .73$) was not different from the error in the homogenous displays ($M = 3.02$, $SD = 1.12$). Moreover, there was a main effect of mean size, $F(1,19) = 27.98$, $MSE = .75$, $p < .001$, $\eta_p^2 = .60$. Participants made more error in the large mean size condition ($M = 3.36$, $SD = 1.03$) compared to the small mean size condition ($M = 2.33$, $SD = 1.64$). Also, the interaction effects did not reach significance, $F(1,19) = 1.58$, $MSE = .39$, $p = .23$, $\eta_p^2 = .08$ (see Figure 18).

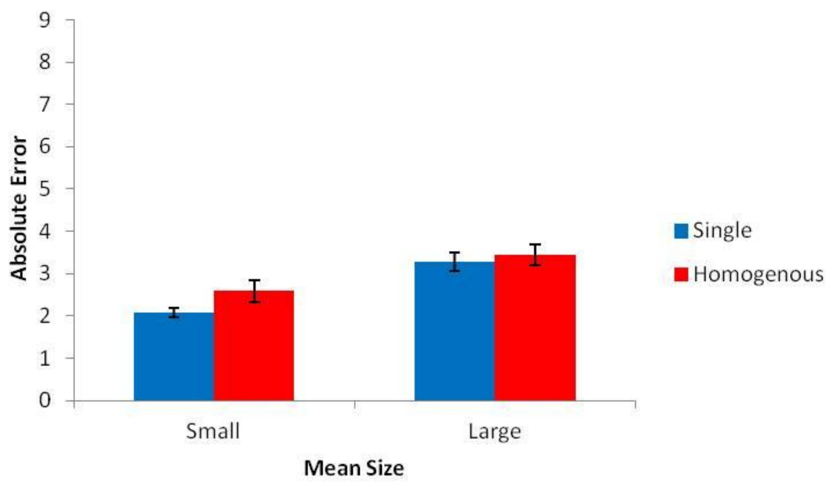


Figure 18. Absolute error in pixels single-homogenous and small-large mean size conditions in Experiment 2A.

Secondly, in order to examine the effects of variance on mean size estimations, we compared the low and high variance conditions in each mean size condition. A 2 (Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. Results again showed a similar pattern. As expected, there was a main effect of variance, $F(1,19) = 17.91$, $MSE = 1.22$, $p < .001$, $\eta_p^2 = .49$. Participants made more error in the high variance condition ($M = 4.76$, $SD = 1.42$) compared to the low variance condition ($M = 3.72$, $SD = .98$). There was no main effect of mean size, $F(1,19) = 2.07$, $MSE = 1.35$, $p = .17$, $\eta_p^2 = .10$, revealing that error in the small mean size condition ($M = 4.05$, $SD = 1.28$) was not different from the error in the large mean size condition ($M = 4.43$, $SD = 1.11$). Finally, there was no interaction effects, $F(1,19) = .80$, $MSE = .63$, $p = .33$, $\eta_p^2 = .04$ (see Figure 19).

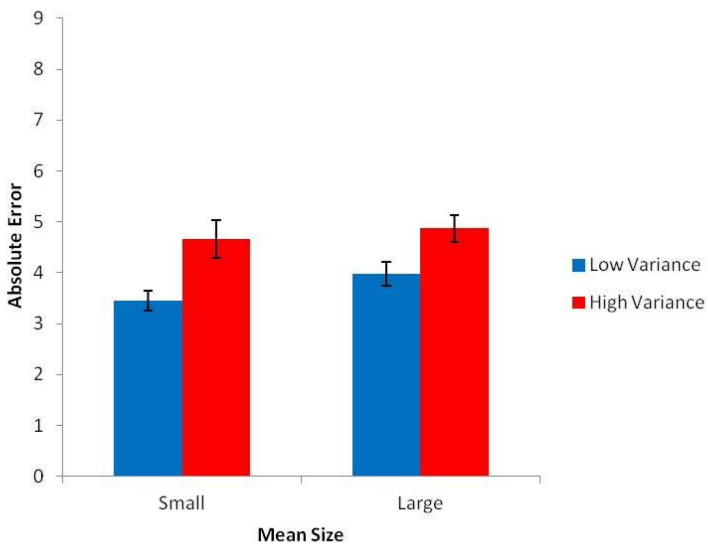


Figure 19. Absolute error in pixels in low-high variance and small-large mean size conditions in Experiment 2A.

Estimation Bias Analysis

We calculated the estimation bias score for each condition similar to the previous experiment. A series of single *t*-tests revealed that only estimation bias score in the high variance condition was significantly different from 1. Similar to Experiment 1, there was an overestimation of the mean size in the high variance condition, $t(19) = 2.80$, $p = .012$ (see Figure 20 & 21).

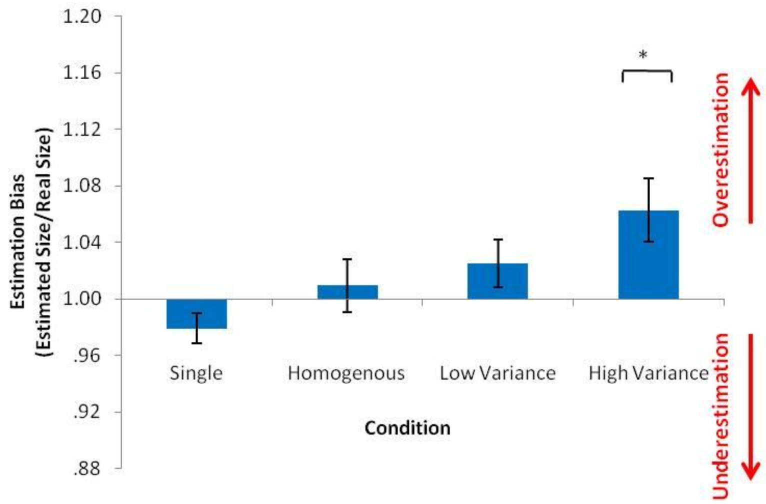


Figure 20. Estimation bias in each condition in Experiment 2A.

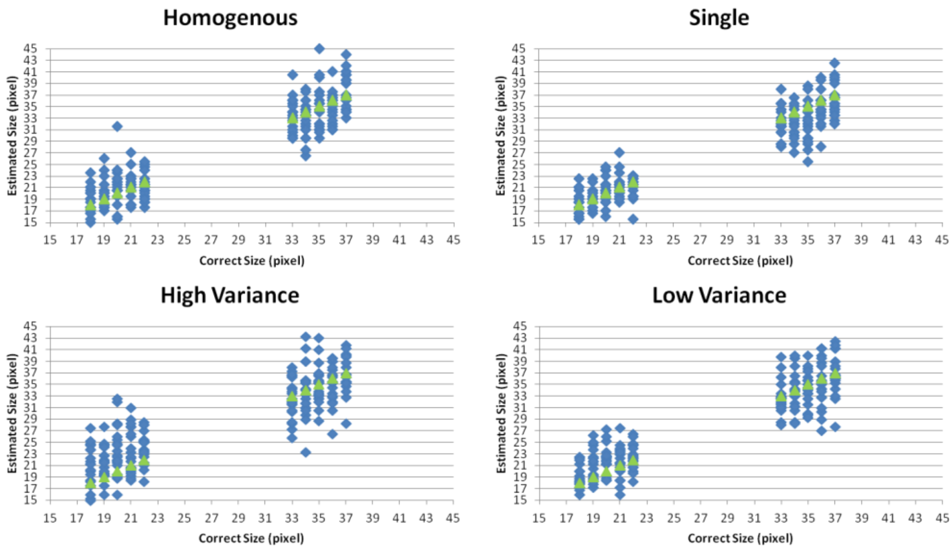


Figure 21. Estimated size (radius in pixels) as a function of correct size in each condition in Experiment 2A (Triangles indicate the correct size).

Real Error Analysis

When compared to 0 by a series of single t-tests, in the high variance condition there was a marginally significant effect, suggesting an overestimation of mean size in the high variance condition, $t(19) = 2.08$, $p = .051$. Also, in the single condition there was a significant effect, suggesting an underestimation of mean size in the single condition, t(19) = -2.63, p = .02 (see Figure 22). The results suggest that when we compress all error participants made within a condition while ignoring the specific trials, we still see an overestimation tendency in the high variance condition. Moreover, the underestimation tendency in the single displays in estimation bias analysis reached significance.

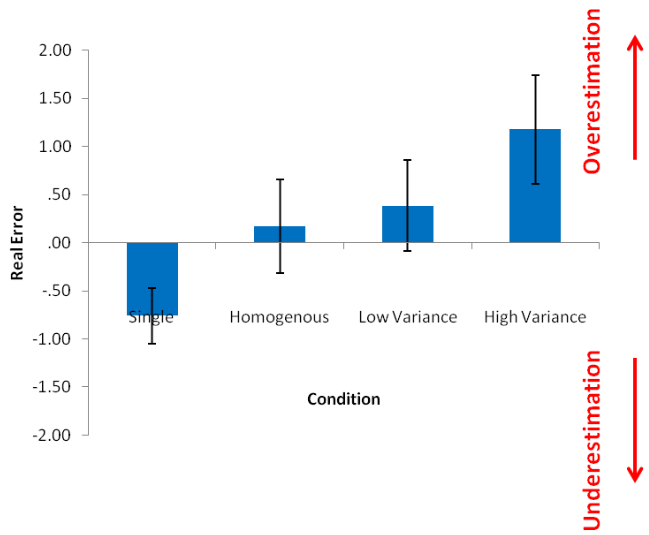


Figure 22. Real error in each condition in Experiment 2A.

Coefficient of Variation

A series of within subjects *t*-test was conducted. First, we compared CV in the homogenous and single conditions. Similar to the first experiment, in the large mean size condition, there were no significant differences between conditions (*M*s = .11, *SD*s = .03), *t*(19) = -.29, *p* = .78, Cohen's *d* = .00, effect size *r* = .00. Same pattern of results was found for the small mean size (*M*s = .14, *SD*s = .04), *t*(19) = .10, *p* = .93, Cohen's *d* = .00, effect size *r* = .00. However, when we compared the high and low variance conditions, we found that in the large mean size CV was higher in the high variance condition (*M* = .16, *SD* = .03) compared to low variance condition (*M* = .13, *SD* = .03), *t*(19) = 5.12, *p* < .001, Cohen's *d* = 1.00, effect size *r* = .45. In the small mean size, although the CV was higher in the high variance condition, it did not reach significance (*M*s = .21 and .19, *SD*s = 04, respectively), *t*(19) = 1.46, *p* = .16, Cohen's *d* = .50, effect size *r* = .24.

Individual Differences Analysis

Visual Working Memory Capacity

Participants Ospan score changed between 28 and 75 with an average of 54.50. In order to examine whether participants' visual working memory capacity was related to their performance, we conducted a series of correlational analysis between Ospan score and absolute error in each condition. Results did not support our expectations. WM capacity does not seem to be related to the performance in any of the conditions (all *r*s < .44, see Figure 23).

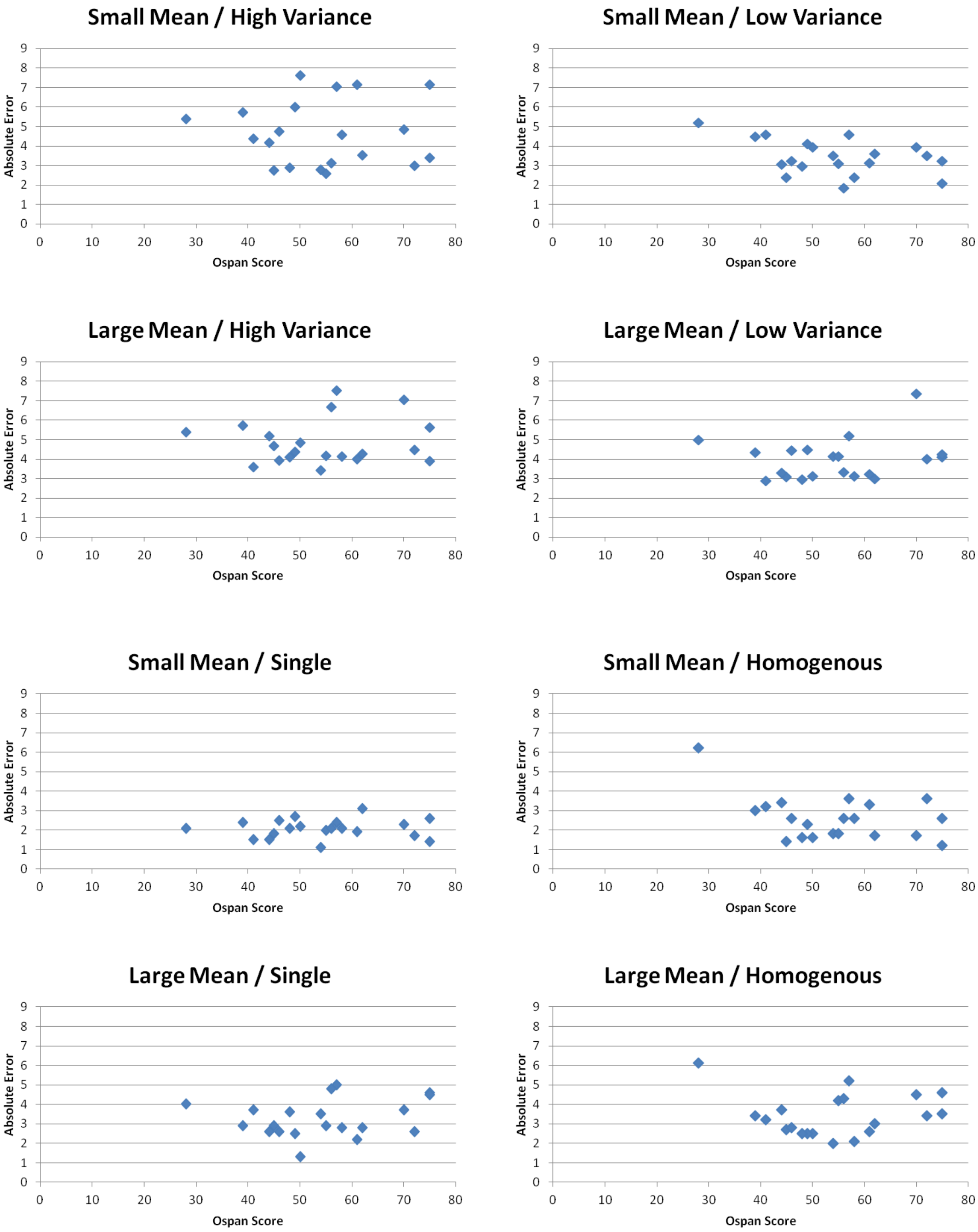


Figure 23. Absolute error as a function of Ospan score in each condition in Experiment 2A.

Sensory Processing Sensitivity and Mental Imagery Preference

In order to examine whether individual differences play a role on the participants' performance, we conducted a series of correlational analysis over 20 participants who filled out the scale measures. There was no significant correlation between participants' estimation bias and their score on HSP scale (all *r*s < .14, all *p*s > .05). Moreover, scores on neither OSIQ object scale nor OSIQ spatial scale were correlated with estimation bias (all *r*s < .28, all *p*s > .05). The results suggest that sensory processing sensitivity or object/spatial imagery did not play a role in the mean size estimation performance.

RT Analysis of the First Key Press

In order to examine the RT effects, after excluding two outlier participants who had RTs 2 or 3 *SD* above the mean, we conducted the analysis over 18 participants. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a main effect of display type, $F(1,17) = 15.67$, $MSE = 11823.17$, $p = .001$, $\eta_p^2 = .48$. RT in the single displays ($M = 690.87$, $SD = 130.64$) was slower than the RT in the homogenous displays ($M = 589.42$, $SD = 132.53$). Also, there was a main effect of mean size, $F(1,17) = 7.53$, $MSE = 6721.14$, $p = .014$, $\eta_p^2 = .31$. Participants were slower in the large mean size condition ($M = 666.67$, $SD = 155.05$) compared to the small mean size condition ($M = 613.63$, $SD = 108.12$). Moreover, there was no interaction effects, $F(1,17) = .0001$, $MSE = 11390.88$, $p = .99$, $\eta_p^2 = .00001$ (see Figure 24).

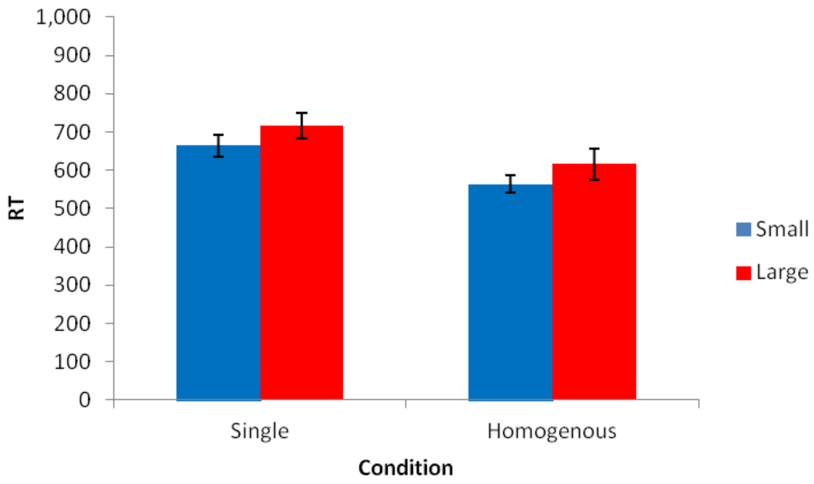


Figure 24. Average RT of the first key press in single-homogenous and small-large mean size conditions in Experiment 2A.

Second, in order to examine the RT differences in different variance conditions, we compared low and high variance conditions in each mean size condition. A 2 (Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a main effect of variance, $F(1,17) = 9.12$, $MSE = 15200.87$, $p = .008$, $\eta_p^2 = .35$. Participants were slower in the high variance condition ($M = 769.76$, $SD = 252.88$) compared to the low variance condition ($M = 681.99$, $SD = 183.06$). However, there was no main effect of mean size, $F(1,17) = .12$, $MSE = 12523.96$, $p = .73$, $\eta_p^2 = .007$. RT in the small mean size condition ($M = 730.44$, $SD = 225.28$) was not different than the RT in the large mean size condition ($M = 721.30$, $SD = 210.66$), and the interaction effect between Variance and Mean Size did not reach significance, $F(1,17) = 1.69$, $MSE = 7789.97$, $p = .21$, $\eta_p^2 = .09$ (see Figure 25).

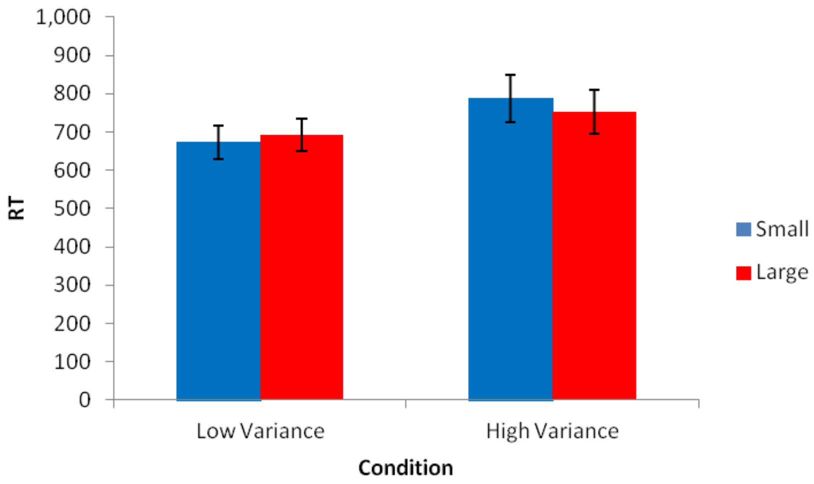


Figure 25. Average RT of the first key press in low-high variance and small-large mean size conditions in Experiment 2A.

Experiment 2B

In Experiment 2A, the orientations of the lines were always horizontal, resulting in collinear displays. We speculated that this characteristic of the displays might bias the estimations. For example, participants might strategically add up the lines while estimating. In order to eliminate this potential bias, in Experiment 2B, we randomly manipulated the orientation of the lines. We believed that this new displays are better in representing the irregularities in the visual displays we encounter in everyday world.

<u>Method</u>

<u>Participants</u>

Twenty one students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation.

<u>Apparatus and Stimuli</u>

The same apparatus and stimuli that were used in Experiment 2A were utilized with one exception: Lines were presented in random orientations rather than horizontally (see Figure 26). While holding constant one side of the lines, we rotated the other side to one of 18 possible locations, each 10° apart. In order to prevent perceptual grouping, no more than two lines with the same orientation were presented within the same set. Also, lines within a set never presented horizontally.

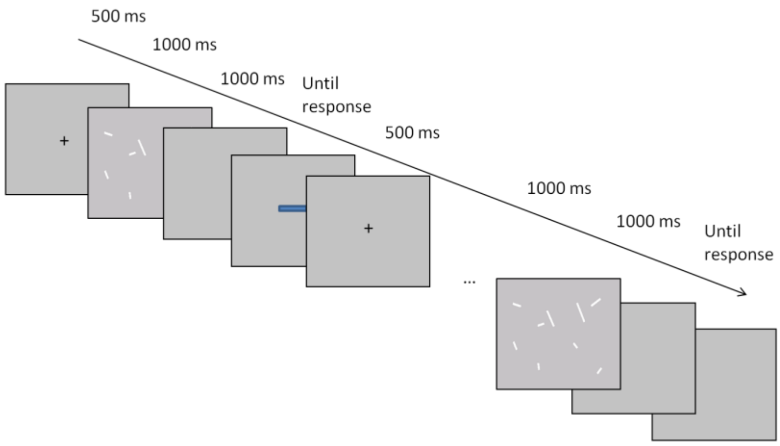


Figure 26. The design of Experiment 2B.

Results

A series of within subjects *t*-tests was conducted to test whether the size of the response circle was equally likely to be larger or smaller than the mean size of the set in each condition (see Table 7). Results showed that in all conditions, the distributions of initial circles were roughly similar (all *t'*s < 1.34 (sign ignored), all *p'*s > .21). Therefore, similar to the previous experiments, the pattern of our findings was not due to the distribution of the initial circles.

Table 7. Mean and Standard Deviations of the Number of Larger and Smaller Initial Radius of the Response Circle than Mean Size of the Set

		Initial Radius			
		Large		Small	
		M	*SD*	*M*	*SD*
Large Mean	Single	5.24	1.26	4.67	1.20
	Homogenous	5.29	1.31	4.48	1.50
	Low Variance	9.71	1.93	9.48	1.78
	High Variance	9.62	2.09	8.67	1.93
Small Mean	Single	5.00	1.55	4.71	1.59
	Homogenous	5.14	1.68	4.52	1.40
	Low Variance	9.67	2.03	9.86	2.13
	High Variance	9.90	2.59	9.62	2.22

Absolute Error Analysis

In order to investigate whether estimation accuracy differed between a single line and 9 same-sized lines (homogenous condition, we compared the single and homogenous displays in each mean size condition. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted.). Interestingly, there was a main effect of display type, $F(1,21) = 8.04$, $MSE = .76$, $p = .01$, $\eta_p^2 = .29$, revealing that participants made more error in the homogenous displays ($M = 3.89$, $SD = 1.20$) compared to the single displays ($M = 3.35$, $SD = .81$). Similar to previous experiments, there was a significant main effect of mean size, $F(1,21) = 26.41$, $MSE = .61$, $p < .001$, $\eta_p^2 = .57$. Participants made more error in the large mean size condition ($M = 4.05$, $SD = 1.15$) compared to the small mean size condition ($M = 3.18$, $SD = .85$). Moreover, the interaction between Display Type and Mean Size variables did not reach significance, $F(1,21) = 4.06$, $MSE = .73$, $p = .06$, $\eta_p^2 = .17$ (see Figure 27).

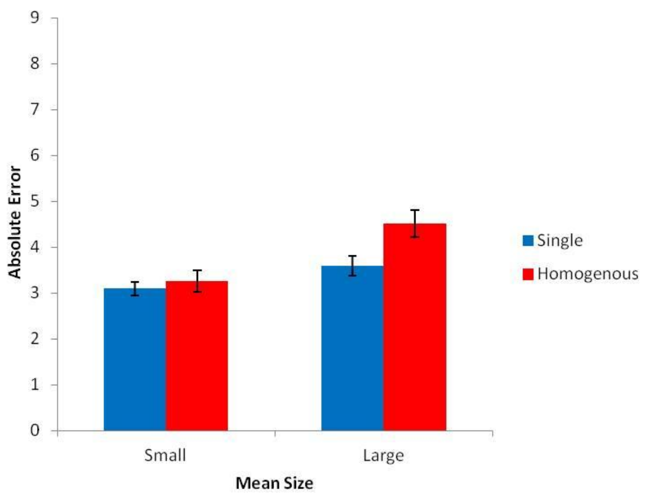


Figure 27. Absolute error in pixels single-homogenous and small-large mean size conditions in Experiment 2B.

Then, in order to examine effects of variance on mean estimations, we compared the low and high variance conditions in each mean size condition. A 2 (Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. Results were similar to the previous two experiments.). There was a main effect of variance, $F(1,21) = 13.35$, $MSE = .84$, $p = .002$, $\eta_p^2 = .40$. Participants made more error in the high variance condition ($M = 5.12$, $SD = 1.29$) compared to the low variance condition ($M = 4.39$, $SD = .97$). However, there was no main effect of mean size, $F(1,21) = 2.35$, $MSE = 1.56$, $p = .14$, $\eta_p^2 = .11$, suggesting that error in the small mean size condition ($M = 4.54$, $SD = 1.22$) was not different from the error in the large mean size condition ($M = 4.96$, $SD = 1.05$). Finally, there was no interaction effects, $F(1,21) = .61$, $MSE = .81$, $p = .45$, $\eta_p^2 = .03$ (see Figure 28).

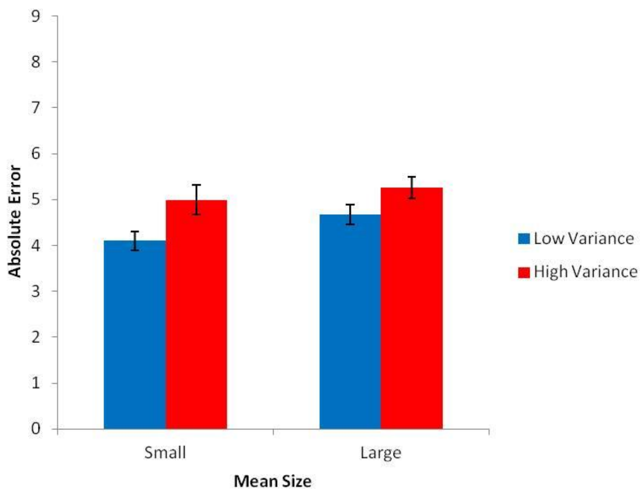


Figure 28. Absolute error in pixels low-high variance and small-large mean size conditions in Experiment 2B.

Estimation Bias Analysis

A series of single t-tests revealed that in all conditions estimation bias score was significantly different from 1, indicating an overestimation of mean size in the single condition, $t(20) = 2.10$, $p = .049$; the homogenous condition, $t(20) = 3.45$, $p = .003$; the low variance condition, $t(20) = 2.96$, $p = .008$; and the high variance condition, $t(20) = 2.28$, $p = .03$ (see Figure 29 & 30). Given the nature of the displays, this finding suggests that irregular displays results in a tendency to overestimate in mean size estimation.

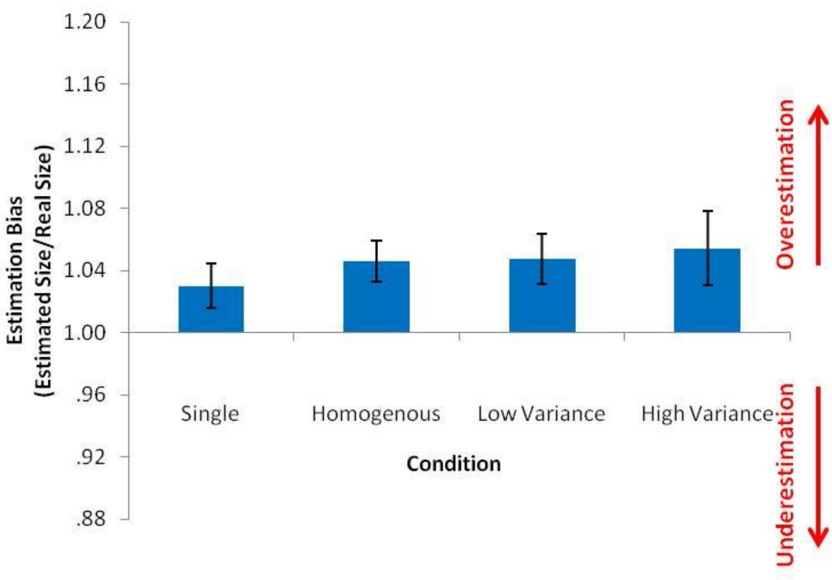


Figure 29. Estimation bias in each condition in Experiment 2B.

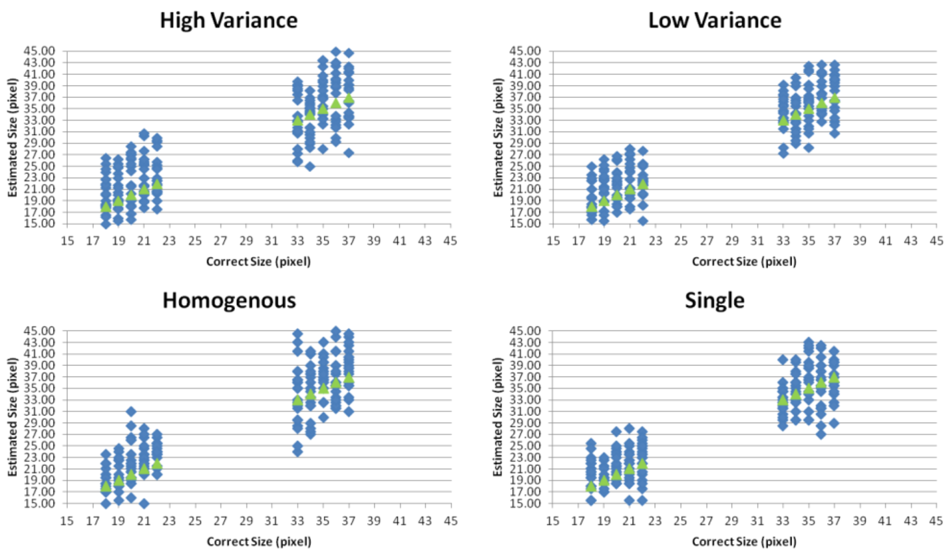


Figure 30. Estimated size (radius in pixels) as a function of correct size in each condition in Experiment 2B (Triangles indicate the correct size).

<u>Real Error Analysis</u>

We compared the real error to 0 by a series of single t-tests. In the high variance condition there was a marginally significant effect, suggesting an overestimation of mean size in the high variance condition, $t(21) = 1.92$, $p = .069$. Also, in the low variance condition there was a significant effect, suggesting an overestimation of mean size in the low variance condition, $t(21) = 2.80$, $p = .011$. Additionally, real in homogenous condition there was a significant effect, suggesting an overestimation of mean size in the homogenous condition, $t(21) = 1.19$, $p = .004$. (see Figure 31). The results suggest that when we compressed all error participants made within a condition while ignoring the specific trials, we still see an overestimation tendency in three of the conditions.

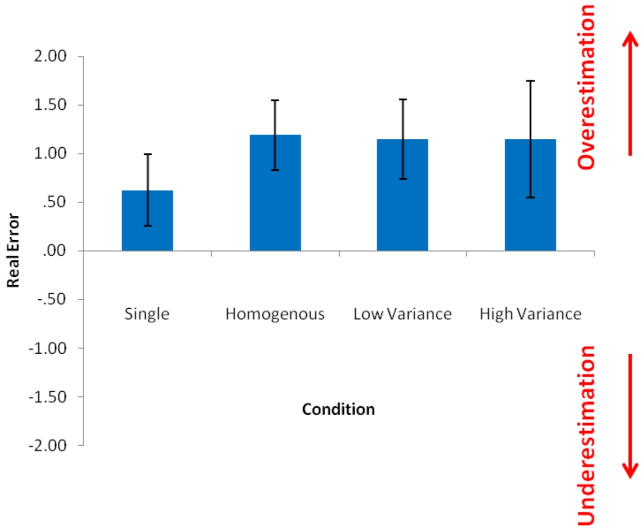


Figure 31. Real error in each condition in Experiment 2B.

Coefficient of Variation

A series of within subjects *t*-test was conducted. First, we compared CV in the homogenous and single conditions. Contrary to previous results, in the large mean size condition, CV was higher in the homogenous condition (*M* = .15, *SD* = .04) compared to the single condition (*M* = .12, *SD* = .03), *t*(20) = 3.59, *p* = .002, Cohen's *d* = .85, effect size *r* = .39. Same pattern of marginal effects was found for the small mean size (*M*s = .20 and .17, *SD*s = .07 and .06, respectively), *t*(20) = 2.07, *p* = .052, Cohen's *d* = .46, effect size *r* = .22. This finding is in line with the lower performance in the homogenous displays in this experiment. Moreover, similar to the previous experiments, in the small mean size we found higher CV in the high variance condition compared to the low variance condition, (*M*s = .25 and .22, *SD*s = .07 and .05, respectively), *t*(20) = 2.13, *p* = .046, Cohen's *d* = .49, effect size *r* = .24. However, in the large mean size, although the CV tend to be higher in the high variance condition (M = .17, SD = .04) compared to the low variance condition (M = .15, SD = .04), difference did not reach level of significance, $t(20) = 1.70$, $p = .10$, Cohen's d = .50, effect size r = .24.

Individual Differences Analysis

Visual Working Memory Capacity

Participants Ospan score changed between 17 and 75 with an average of 54. In order to examine whether participants' visual working memory capacity was related to their performance, we conducted a series of correlational analysis between Ospan score and absolute error in each condition. We did not find any significant correlations; therefore, in contrast to our expectations, WM capacity does not seem to be related to the performance in any of the conditions (all rs <.38, see Figure 32).

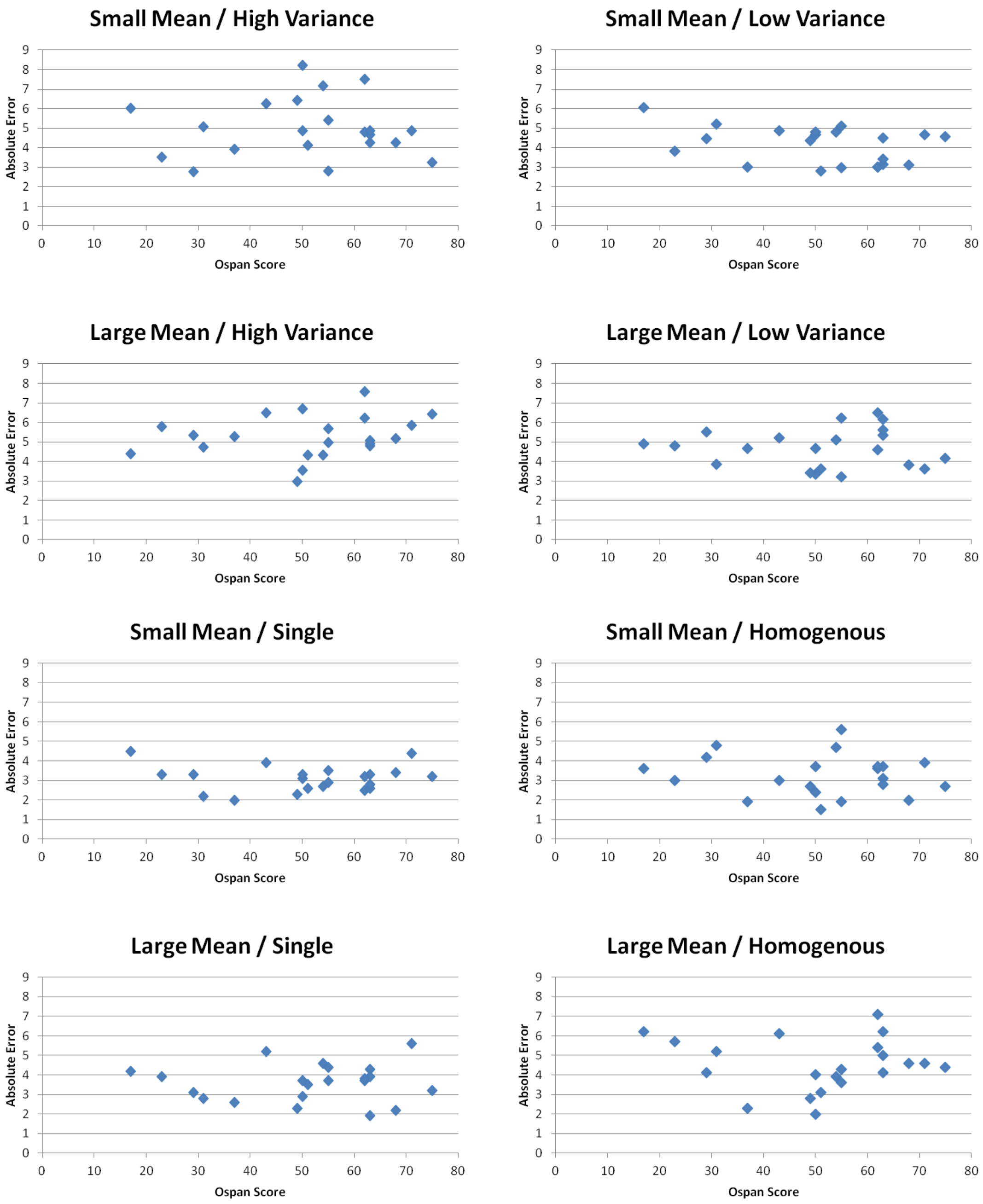


Figure 32. Absolute error as a function of Ospan score in each condition in Experiment 2B.

Sensory Processing Sensitivity and Mental Imagery Preference

In order to examine whether individual differences play a role on the participants' performance, we conducted a series of correlational analysis over 19 participants who filled out the scale measures. There was no significant correlation between participants' estimation bias and their score on HSP scale (all *r*s < .182, all *p*s > .05). However, scores on OSIQ object scale and OSIQ spatial scale were positively correlated with estimation bias only in single displays (*r*s > .48, *p*s < .05). The results suggest that sensory processing sensitivity or object/spatial imagery did not play a significant role in the mean size estimation performance.

RT Analysis of the First Key Press

After four outlier participants who had RTs 2 or 3 *SD* above the mean, we conducted the analysis over 17 participants. A 2 (Display Type: Single vs. Homogenous) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a marginal main effect of display type, $F(1,16) = 4.27$, $MSE = 18857.01$, $p = .055$, $\eta_p^2 = .21$. RT in the single displays ($M = 653.61$, $SD = 195.76$) was slower than the RT in the homogenous displays ($M = 584.75$, $SD = 142.165$). However, there was no main effect of mean size, $F(1,16) = .24$, $MSE = 18074.57$, $p = .63$, $\eta_p^2 = .02$. RT in the large mean size condition ($M = 627.09$, $SEM = 146.67$) was not different than the RT in the small mean size condition ($M = 611.27$, $SD = 191.26$). Moreover, there was no interaction effects, $F(1,16) = 1.23$, $MSE = 13400.16$, $p = .28$, $\eta_p^2 = .07$. Second, in order to examine the RT differences in different variance conditions, we compared low and high variance conditions in each mean size condition. A 2

(Variance Type: Low vs. High) X 2 (Mean Size: Small vs. Large) within subjects ANOVA was conducted. There was a main effect of variance, $F(1,16) = 6.36$, MSE = 10950.88, $p = .023$, $\eta_p^2 = .28$. Participants were slower in the high variance condition (M = 683.41, SD = 192.69) compared to the low variance condition (M = 619.42, SD = 181.29). However, there was no main effect of mean size, $F(1,16) = .70$, MSE = 11899.25, $p = .42$, $\eta_p^2 = .04$. RT in the small mean size condition (M = 640.34, SD = 168.89) was not different than the RT in the large mean size condition (M = 662.48, SD = 205.10), and the interaction effect between Variance and Mean Size did not reach significance, $F(1,16) = 2.08$, MSE = 12317.66, $p = .17$, $\eta_p^2 = .12$.

Discussion

The present set of experiments was conducted in order to examine the mechanisms that systematically affect ensemble representations, specifically mean size estimations. First, we tested whether individuals were able to compute mean size of a given set. Second, we investigated whether these estimations were biased as a function of the within-set variance. Moreover, individual characteristics which were potential predictors of performance on the formation of ensemble statistics (visual working memory capacity, sensory processing sensitivity and visual imagery ability) were investigated. We also tested whether the diameter of the circles was a strong candidate for the unit of averaging in mean size estimations. Finally, we conducted additional analysis on mean size estimations in order to suggest directions for further research.

In the first two experiments we demonstrated that performance in the single and homogenous displays were similar and people were highly accurate in their performance with errors being less than .15° of visual angle. The variability of participants' responses in these conditions was also similar. These two conditions served more of a control for our experiment. The high accuracy favors the automaticity of ensemble representations (Ariely, 2001; Chong & Treisman, 2003, 2005a-b) because participants represented the items readily and accurately. Similarly, Chong and Treisman (2003) reported that performance on homogenous and single displays were similar, diameter difference being around 6%-8% for 75% accuracy in mean size discriminations. In Experiment 3; however, we demonstrated that performance in the homogeneous displays was worse and more variable. However, we believe that this drop in performance was due to the fact that in these displays, contrary to the displays in the first two experiments, there were two variables changing dimensions. Both size/length and orientation of the lines were changing. This random manipulation resulted in irregular displays making it hard to identify the fact that the lines were of equal length. Even in the single displays, participants had to rotate the line mentally to estimate its size while responding, leading to an increase in the error. Interestingly, in the single displays of all experiments, participants were slower in making their first response.

In all three experiments, we also demonstrated that the higher within-set variance resulted in poorer performance with errors being less than .26°and slower response times in making the first response. Also, participants' responses revealed more variability in the high variance condition. Therefore, it seems that in the context of the mean size estimations, by being ultimately influenced by the nature of the set, people were not able to correct for the variations in displays. Participants seemed to take their time to adjust the size in their minds before making any response (Rosenbaum & Kornblum, 1982). Our findings build on Chong and Treisman's

(2003)'s suggestion of the significance of within-set distributions. They are unique in the sense that we were able to demonstrate that the nature of the displays biased the computation the mean size. Recently, Sims, Jacobs and Knill (2012) reported that the distribution of other items within a set affects the representation of an individual item with error being higher in high variance. In the present set of experiments, we were able to extend these previous findings to mean size estimations.

More significantly, in all three experiments, we demonstrated an overestimation tendency in the high variance conditions. We speculated that higher variance increases irregularity of a set, leading more variable and systematically biased estimations. The fact that there was an overestimation tendency in all conditions of Experiment 2B supported this reasoning.

The tendency to overestimate the general summary of irregular items was demonstrated by earlier studies in different contexts. For instance, Charras and Lupiáñez (2009) demonstrated that in a line-length comparison task, that sum of two asymmetrically bisected parts of a line were perceived as greater than the sum of two symmetrically bisected parts. Later, Charras, Bord and Lupiáñez (2012) extended this finding to a number comparison task, reporting that sum of repeated numbers were underestimated while sum of different numbers were overestimated. We were able to demonstrate that, this overestimation tendency in summation is also evident in averaging. There is also prior evidence that addition and averaging is influenced by similar mechanisms (Van Opstal et al., 2011).

Moreover, in all three experiments, we demonstrated that performance in the variance conditions was not affected by the mean size. However, in the homogenous and single conditions, larger mean size resulted in more error and slower response times. This finding suggests that when participants mentally computed the mean size (in variance conditions), their performance was not affected by how large/small their estimations were. However, when participants saw the actual size of the circles (in homogenous and single conditions), they made more error while reproducing the large circles. Also, slower response times in estimating the size of large circles might indicate that participants took longer time to adjust their responses in their minds before making the first response. Since large numbers open the possibility of higher dispersion, the representation of the size of larger circles may be prone to more error.

CHAPTER 3

VARIANCE ESTIMATIONS AS AN ENSEMBLE REPRESENTATION

Although average size is the most common and currently available descriptive statistics used in the ensemble representation studies, researchers had previously investigated people’s variance estimation abilities (e.g., Beach & Scopp, 1968; Hofstatter, 1939; Lovie & Lovie, 1976). Ability to accurately estimate variance is meaningful both practically and theoretically. Such estimation provides a more accurate representation of a given visual display. Representing variance within occurrences of a repetitive event is important for one's ability to make accurate predictions about future occurrences. Hence, we believe that it is meaningful to investigate the effects of variance on ensemble representations.

One of the earlier works on estimation of variance was conducted by Hofstatter (1939, as cited in Peterson, & Beach, 1967) who asked participants to estimate the variability in the lengths of tied bundles of sticks. A positive correlation between real and estimated variability was reported. Also, he found that as the mean length increased, variability judgments decreased. In other words, observers seemed to estimate variability of the sticks but their judgments were influenced by the magnitude of the mean length of the bundles.

Similarly, Beach and Scopp (1968) demonstrated that variance estimations were systematically biased by other parameters of a given set. They presented observers index cards; on each card there were two-digit numbers. They manipulated the mean and variance of the numbers written on the sets. They presented participants with two cards at a time (20 cards for a trial), one at their right hand side and another at their left hand side. The task was, after all 20 cards were presented, to judge which population of cards had larger variance and estimate its ratio. Results showed that participants were inaccurate in their judgments because they underestimated the weight of deviant sample and estimated large numbers as less deviant then small numbers. Their judgments were also influenced by the magnitude of the mean, similar to Hofstatter's (1939, as cited in Peterson, & Beach, 1967) suggestion.

Later, the relationship between mean and variance estimations of a given set was examined by Lovie and Lovie (1976). They presented participants with a list of numbers and asked them to report first the mean then the variance of the list. They found that participants were accurate in mean estimations. But there was a trend that variance of the sample affected the mean estimations. Mean estimations were more accurate in the low variance condition. People were worse at estimating variance than means, and mean estimations were influenced by the magnitude of the variance. They also reported that variance estimation performance improved significantly over trials only in the low variance condition. Interestingly, size of the list did not affect the performance. They also examined the effect of different mean conditions (low versus high) on estimation of variances. Similar to the earlier work by Hofstatter (1939, as cited in Peterson, & Beach, 1967), they found that higher variance estimates were found when the mean size was lower. However, they did not find any interaction between mean and variance sizes on variance estimates. That is, despite different mean sizes, variance was still estimated to be higher in high variance condition and lower in the low variance condition. They speculated that this finding may result from the instructions because they explicitly told participants to estimate variances without taking mean into account.

As seen above, there has been little research focusing directly on perception of variability, yielding inconsistent findings. Pollard (1984) reviewed the relevant studies and reported that people in fact seem to be sensitive to sample variability. However, making clear conclusions is not possible due to methodological restrictions. There has been even less research on whether there is a systematic bias in these types of variance judgments. To our knowledge only Kareev, Arnon and Horwitz-Zeliger (2002) examined whether mean and variance estimations systematically influence each other in a controlled study. In their Experiment 1, they presented participants with one set of items of two types: either plastics disks which were marked with one or two letters or paper cylinders which were colored up to a certain height. Later, they presented another two sets of items from the same type: one set has the same variability with the original set while the other has either less or more variability compared to the original set. For instance, if participants saw a set of plastic disks first, later the same set was presented along with a new set of plastic disks. The new set has either lower or higher variability. The task was to judge which one of the new groups was more similar to the first one. Results showed that the identical group was chosen more when it was given with a larger variable group (.66 of the cases) than with a smaller one (.49 of the cases). Therefore, they claimed that people estimate population variability smaller than it really is. These findings were in line with the idea that perception of variability is “downward attenuated” (Kareev et al., 2002, p. 287). That is, individuals are prone to consider the world to be regular or less variable. The fundamental attribution error (Gilbert & Malone, 1995; Ross, 1977) phenomenon can be an example of this tendency. People exhibit a tendency to perceive others' behavior as less variable than it is. More importantly, these findings indicated that people were sensitive to variability differences within groups. Another key finding from other experiments by Kareev et al. was that participants with smaller WM capacity reported lower variability than those with larger capacity. They suggested that this may be evidence that people attend to use sub-samples without making any correction for sample values before judging population. People may need to subsample population in order to estimate the larger population, mostly due to time or memory constraints.

The finding that observers with smaller WM reported lower variability of a given set of items can be seen as in line with the subsampling hypothesis. Furthermore, Kareev, Liberman, and Lev (1997) reported that people were more accurate and faster in estimating the correlation between two variables if they were given a small sample or if they had a small WM capacity. Also, these people estimated greater correlations. It calls further research to investigate the WM capacity differences in average size estimation judgments. More significantly, this opens a more interesting area of research. Individual differences approach can be used to investigate the ability to compute ensemble statistics.

Experiment 3A

Most previous research which examined variance estimations as ensemble statistics used numbers as stimuli (e.g., Lovie & Lovie, 1976). Although results generally showed an indication of accurate variance estimations (for an exception, Beach & Scopp, 1968), there are other findings on the representation of variance lower than it is (Kareev et al., 2002). The nature of the variance estimations using other more controlled stimuli and methodologies which are more sensitive to errors can be beneficial. In Experiment 3, we directly tested if people are able to accurately estimate variance of a given set of circles. We hypothesized that as the variance difference between the distractor and the target sets increases, discrimination ability should improve. Also, reflecting on previous findings, we hypothesized that accuracy should be higher when the distractor set has higher than lower variance than the target set.

Method

Participants

Twenty four students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation.

Variance Estimation Task

Each trial started with a fixation cross presented at the center of the screen for 500 ms (see Figure 33). Following the cross, a set of 9 circles appeared for 1000 ms followed by a 1000 ms interval, after which two different sets of circles appeared at each side of the screen for 1000 ms. One of these displays was the same as the first set but the individual circles was positioned within a different configuration. The other display was its pair which had the same mean but had either higher or lower variance. Then, a blank screen was presented for 2000 ms. Participants were able to make their responses either when two sets were on the screen or when the blank screen was present. There were 64 trials in total.

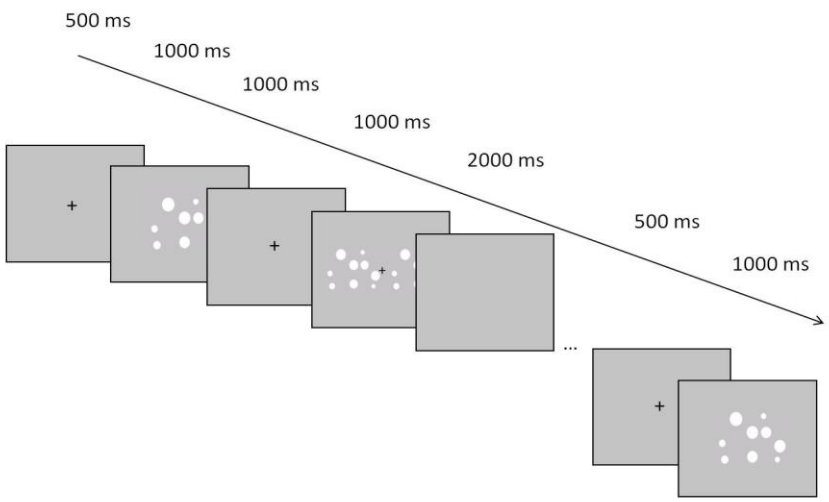


Figure 33. The time-line of the variance estimation task in Experiment 3A.

Each display was presented as the target only once. The task was to judge which of the two displays was more similar to the previous set. Participants pressed the keys ‘k’ and ‘d’ to indicate either the right or the left display. Conditions were presented in a mixed fashion. All displays were randomly presented with a constraint that target sets with the same mean size or variance could not appear consecutively more than three times. Also, the side of the screen the target and distractor sets was displayed was counterbalanced in terms of both the mean size and the variance of the set. There were 14 practice trials at the beginning of the task in which accuracy and RT feedback was provided.

One hundred and twenty eight displays, each containing 9 black circles appeared on a gray background were created. Sixty four displays served as target sets; half had smaller (40 pixels) and the other half had larger mean sizes (60 pixels).

Each display was paired with a distractor which had the same mean size as the target but the variance was either higher or lower (see Table 8).

Table 8. Trial Information for Experiment 3A (diameter in pixel)

M	*SD* (Target)	*SD* (distractor) Low Variance	# of Trials	*SD* (distractor) High Variance	# of Trials	Total # of Trials
40	12	10	4	20	4	8
	14	10	4	20	4	8
	16	10	4	20	4	8
	18	10	4	20	4	8
60	12	10	4	20	4	8
	14	10	4	20	4	8
	16	10	4	20	4	8
	18	10	4	20	4	8
Total # of Trials			32		32	64

The diameter of each individual circle was randomly selected from a normal distribution spanning the interval of 20 to 80 pixels (1° – 4°). None of the individual circles had the size of the mean. Also, there was no repetition size within a set.

The position of the circles in the targe displays was randomly determined within a 3 x 4 grid (12.45° x 19°); with a jitter of +/- 5 pixels to minimize regularities. Also, in order to avoid occlusion of stimuli with the edges of the screen, the matrix was positioned with a restriction of 50 pixels (2.5°) away from the top and bottom of the screen.

The position of the circles in the target-distractor displays was randomly determined within two 3 x 4 grids (11.4° x 19°); with a jitter of +/- 5 pixels to minimize regularities. Two matrices were separated by 42 pixels (2.1°) from each other and 50 pixels (2.5°) from the each side of the screen.

Procedure

Participants first completed the Variance Estimation Task, and then filled out the demographic form. Finally, they were debriefed. All stimuli, timing operations and data collection were controlled by a PC running E-prime software (Psychology Research Tools, Inc).

Results

Accuracy and Reaction Time Analysis

In order to examine whether accuracy was higher when the distractor set had higher variance than the target set, we conducted a within subjects *t*-test. Results showed that there was no difference between high variance (*M* = 60.68, *SD* = 12.05) and low variance (*M* = 63.54, *SD* = 13.09) distractors in terms of accuracy, *t*(23) = -.89, *p* = .38, Cohen's *d* = -.23, effect size *r* = -.11. The RT data mimicked the same pattern of results, *t*(23) = .83, *p* = .42, Cohen's *d* = .07, effect size *r* = .04.

Amount of Difference in terms of Variance

To examine the impact of the variance difference between distractor and target sets, we compared accuracy in each amount of variance difference within each distractor type. A 4 (Amount of Variance Difference: 2, 4, 6, 8) X 2 (Distractor Type: High vs. Low) within subjects ANOVA was conducted. Similar to findings above, there was no main effect of distractor type, $F(1,23) = .79$, $MSE = 500.88$, $p = .38$, $\eta_p^2 = .03$.

However, there was a main effect of amount of variance difference, $F(3,69) = 7.51$, $MSE = 242.82$, $p < .001$, $\eta_p^2 = .25$. Post-hoc comparisons showed that when the variance difference was 2 pixels between target and the distractor sets ($M = 53.39$, $SD = 20.39$), accuracy was lowest compared to all other conditions (*M*s = 62.76, 64.58 and 67.71; *SD*s = 20.28, 16.87 and 19.95; *ps* = .031, .002 and .002, Bonferroni-corrected for 4, 6 and 8 pixels respectively). There was no interaction between two variables, $F(3,69) = .40$, $MSE = 348.87$, $p = .75$, $\eta_p^2 = .02$ (see Figure 34).

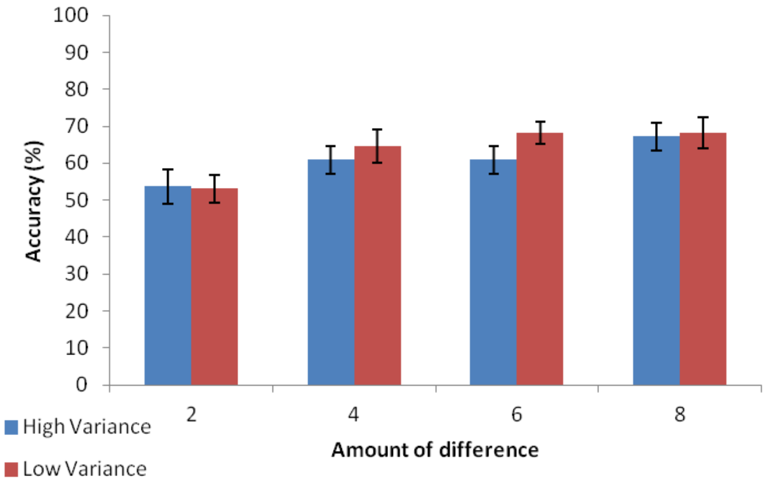


Figure 34. Percentage accuracy in each amount of variance difference between target and distractor set within each variance condition in Experiment 3A.

Experiment 3B

In Experiment 3B, we utilized longer exposure durations, thus allowing participants to view target sets longer. We increased the target display exposure duration from 1000 to 2000 ms and the target and distractor display exposure duration from 1000 to 3000 ms.

Method

Participants

Twenty students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation.

Results

Accuracy and Reaction Time Analysis

As in Experiment 3A, there was no difference between the high variance and low variance conditions in neither accuracy ($t(20) = 1.05$, $p = .31$, Cohen's $d = .024$, effect size $r = .12$) or RT ($t(19) = -1.65$, $p = .12$, Cohen's $d = -.09$, effect size $r = -.05$). For purposes of brevity, we do not report descriptives for these conditions.

Amount of Difference in terms of Variance

To determine the impact of the amount of variance difference on performance, a 4 (Amount of Variance Difference: 2, 4, 6, 8) X 2 (Distractor Type: High vs. Low) within subjects ANOVA was conducted. As in Experiment 3A, there was no main effect of distractor type, $F(1,19) = 1.11$, $MSE = 427.43$, $p = .31$, $\eta_p^2 = .06$ and no significant interaction, $F(3,57) = .61$, $MSE = 329.43$, $p = .62$, $\eta_p^2 = .03$. However, there was a main effect of amount of variance difference, $F(3,57) = 5.43$, $MSE =$

263.30, p = .002, η_p^2 = .22. Post-hoc comparisons showed that when the variance difference was 2 pixels between target and the distractor sets (M = 59.06, SD = 16.82), accuracy was lower compared to a difference of 6 pixels (M = 72.19, SD = 23.44, p = .039, Bonferroni-corrected) or 8 pixels (M = 70.00, SD = 23.74, p = .038, Bonferroni-corrected); but it was not different from the 4 pixels difference (M = 63.75, SD = 17.81), all ps < 1.00, Bonferroni-corrected) (see Figure 35). Therefore, we replicated Experiment 3A results.

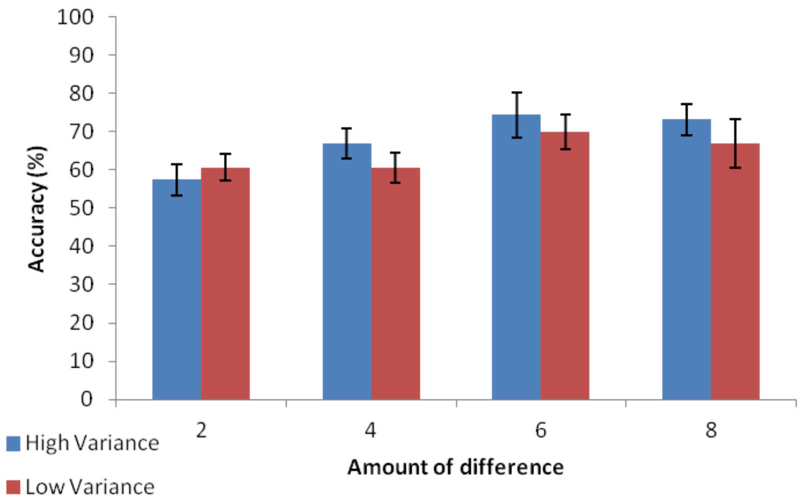


Figure 35. Percentage accuracy in each amount of variance difference between target and distractor set within each variance condition in Experiment 3B.

Experiment 3C

Although the accuracy levels increased in Experiment 3C with longer exposure duration, there still was no performance difference between different distractor types. We speculated that this null finding may be partly due to the instructions we used. It is possible that participants did not understand what was meant by “similarity” (*benzerlik* in Turkish) between sets. Although we told participants to ignore the specific positions of individual circles within each set, they might have relied on this information unknowingly. To overcome this problem, in Experiment 3C, we used more detailed explanation of the variance concept, provided more examples and increased the practice duration (see Appendix D for instructions). We emphasized that the mean size between targets and distractor sets would always be equal and that participants judgments should be based on the similarity of the variance of these displays. As part of Experiment 3C, we replicated the procedures from 3A (N = 22) & 3B (N = 24). For purposes of brevity, we do not report detailed analyses on these findings which essentially replicated the earlier reported null findings. For neither encoding duration, did the performance on different variance conditions differed in terms of accuracy, *t*s < 1.22, *p*s > .24. RT results showed the same pattern in both short exposure durations ($t = 1.39$, $p = .18$) while a different pattern emerged for long durations.[6] As in the previous two experiments, the amount of variance difference had an impact on performance. As the variance difference between studied and tested displays increased, performance improved.

Experiment 4A

One of the predictions in Experiment 3 was that the representation of variance would be attenuated; however, there was no evidence in favor of "downward attenuation" of variance. We speculated that the null effects observed might have been due to the particular task we used in Experiment 3. In that task, we had presented participants first with a target set which was then followed by two sets, one

[6]In the long exposure duration, participants were slower when the distractor has lower than higher variance than the target, $t(23) = -2.64$, $p = .02$, Cohen's $d = -.23$, effect size $r = -.12$.

of which was the same as the target set while the other was a different set with either higher or lower variance. The notion of “downward attenuation” suggests that the target set would be represented with a lower variance than it actually has. However, it is possible that we washed out this effect by presenting the target set again at test. Thus, in Experiment 4, we presented the target initially and then after a delay we presented the test stimuli for comparison. Also, rather than increasing the exposure duration of the visual display, in this experiment we reduced the exposure duration. This latter manipulation was in line with findings that people can in 50 ms form ensemble representation of viewed displays (Chong & Treisman, 2003) and it was targeted to minimize reliance on specific clues form the displays. In Experiment 3, in all phases, about 91% of participants reported focusing on the specific items in the viewed displays.

Method

Participants

Twenty nine students taking the introductory psychology course at Boğaziçi University participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation. Only the participants who performed 80% and above in the practice trials were included in the analysis. There were 21 participants who reached this criterion.

Variance Estimation Task

Each trial started with a fixation cross presented at the center of the screen for 500 ms (see Figure 36). A short bleep was presented through the earphones when the fixation cross was on the screen. Following the cross, a set of 9 circles appeared for 200ms followed by a 900 ms interval, after which a probe set of circles appeared at center of the screen for 3000 ms. Then, a blank screen was presented for 1000ms. Participants were able to make their responses either when two sets are on the screen or when the blank screen was present.

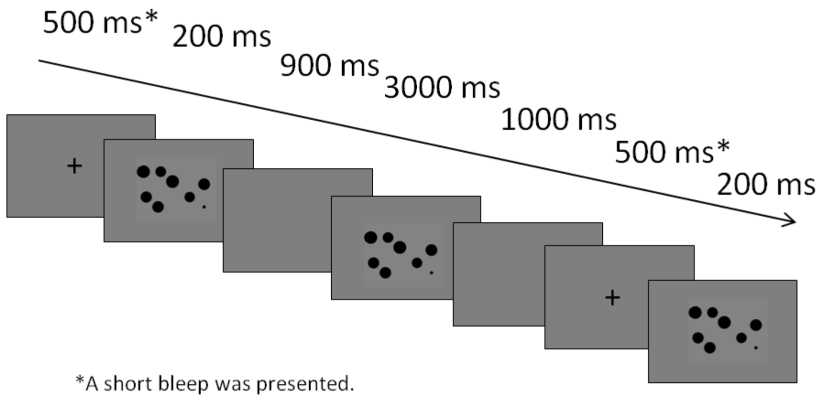


Figure 36. The time-line of the variance estimation task in Experiment 4A.

Each display was presented as the target only once. The task was to judge whether the probe set had the same variance as the target set. Participants pressed the ‘k’ and ‘d’ keys to indicate either the same or different judgment, respectively. All displays were randomly presented with a constraint that target sets with the same mean size or variance could not appear consecutively more than three times.

At the beginning of the task, there was a long and detailed training on the concept of variance, accompanied by several examples using set of numbers and circles (see Appendix E for the instructions). There were 10 practice trials at the beginning of the task which provide accuracy and RT of the response as feedback. In the practice trials, participants were presented with two sets of circles at the each side of the screen. These two sets had the same mean size but different variance. The task was to judge which set had the higher variance. Participants responded using left and right arrow keys to indicate the set placed in the corresponding side of the screen. They had to complete the practice trials with at least 70% accuracy in order to move to the experimental trials. If a participant was unable to reach the accuracy cut-off, a message appeared on the screen informing participants that they had to repeat the practice trials.

One hundred and sixty displays containing 9 black circles appeared on a gray background were created. Since there were 80 trials in total, 80 displays served as target sets. Remaining sets served as the probe displays. In the same condition, there were 40 displays which had the same mean and variance as the target set but were composed of different individual circles. In the different conditions, there were 40 displays. In the High Variance condition (20 trials), the probe set had the same mean size but higher variance than the target set. On the other hand, in the Low Variance condition (20 trials), the probe set had the same mean but lower variance than the target set (see Table 9).

Table 9. Standard Deviations of Each Set in Experiment 4A (in pixel)

Amount of Difference	Target	Same (50% of trials)	Different High Variance (25% of trials)	Different Low Variance (25% of trials)
4	18	18	22	14
6	18	18	24	12
8	18	18	26	10
10	18	18	28	8
4	20	20	24	16
6	20	20	26	14
8	20	20	28	12
10	20	20	30	10
4	22	22	26	18
6	22	22	28	16
8	22	22	30	14
10	22	22	32	12
4	24	24	28	20
6	24	24	30	18
8	24	24	32	16
10	24	24	34	14
4	26	26	30	22
6	26	26	32	20
8	26	26	34	18
10	26	26	36	16

The diameter of each individual circle was randomly selected from a normal distribution spanning the interval of 15 to 95 pixels (0.75° – 4.75°). The position of the circles in the displays was randomly determined within a 3 x 3 grid (18° x 18°); with a jitter of +/- 10 pixels to minimize regularities. The matrix was positioned 140 pixels (7°) away from the right and left side of the screen. Also, in order to avoid occlusion of stimuli with the edges of the screen, the matrix was positioned with a restriction of 60 pixels (3°) away from the top and bottom of the screen.

Procedure

Participants first completed the Variance Estimation Task, and then filled out the demographic form. Finally, they were debriefed.

Results

Accuracy and Reaction Time Analysis

In order to examine whether performance was different in each condition, we compared percentage of accuracy with a within subjects ANOVA with variance condition (same, high and low) as the within subject variable and accuracy as the dependent variable. There was an effect of condition, $F(2,40) = 5.53$, $MSE = 154.62$, $p = .008$, $\eta_p^2 = .22$. Post-hoc comparisons showed that accuracy was lower in the high variance condition ($M = 54.52$, $SD = 15.40$) compared to the low variance condition ($M = 67.14$, $SD = 12.41$, $p = .025$, Bonferroni-corrected), but not different from same condition ($M = 59.17$, $SD = 8.56$), all ps $< .67$, Bonferroni-corrected. There was no other significant difference between conditions (see Figure 37).

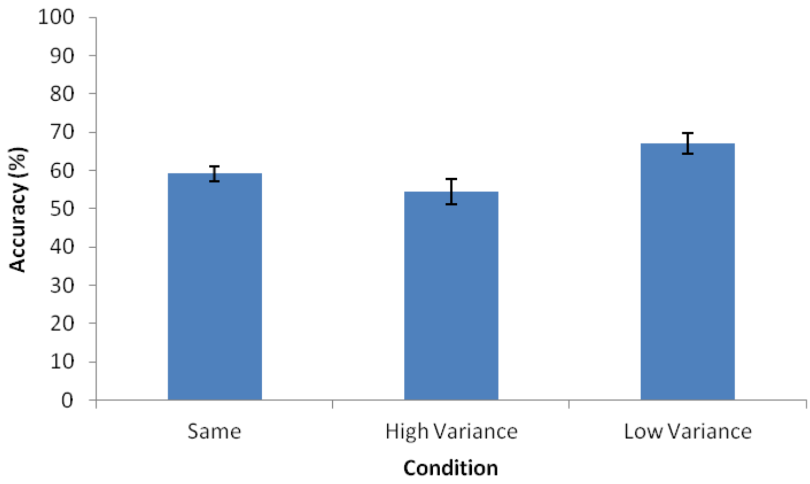


Figure 37. Percentage accuracy in each condition in Experiment 4A.

RT data was fully consistent with the accuracy data. A within subjects ANOVA with condition (same, high, low) as the within subject variable and accuracy as the dependent variable yielded a main effect of condition, $F(2,40) = 6.13$, MSE = 17066.78, $p = .005$, $\eta_p^2 = .24$. Post-hoc comparisons showed that RT was slowest in the high variance condition (M = 1399.78, SD = 187.734) compared to both same condition (M = 1300.23, SD = 283.55, $p = .044$, Bonferroni-corrected) and low variance condition (M = 1263.30, SD = 241.07, $p = .014$, Bonferroni-corrected), all ps < 1.00, Bonferroni-corrected. There was no other significant difference between conditions (see Figure 38).

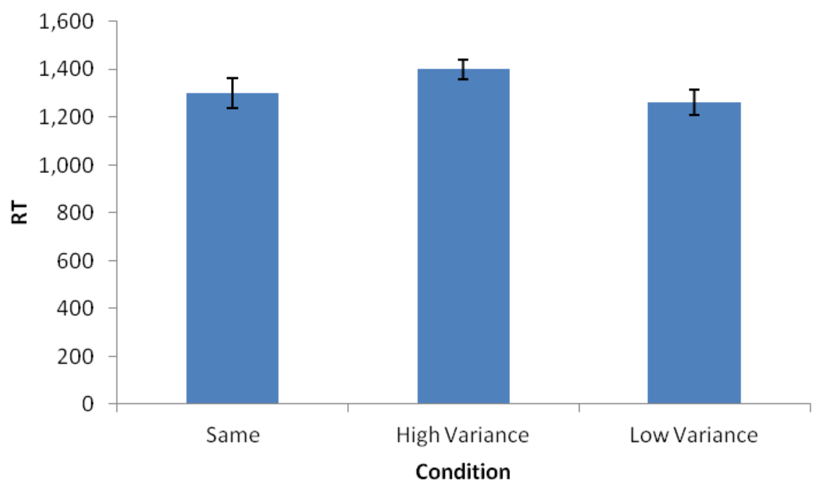


Figure 38. Average correct RT in each condition in Experiment 4A.

Amount of Difference in Terms of Variance

In order to examine whether as the variance difference between the target and probe sets increased, performance improved, we compared accuracy in each amount of variance difference within each different condition. A 4 (Amount of Variance Difference (2, 4, 6, 8) X 2 (Condition: High vs. Low) within subjects ANOVA was conducted. There was a main effect of variance condition, $F(1,20) = 8.60$, $MSE = 778.10$, $p = .008$, $\eta_p^2 = .30$. Post-hoc comparisons showed that accuracy was lower in high variance condition ($M = 54.52$, $SD = 15.40$) compared to different low variance condition ($M = 67.14$, $SD = 12.41$), $p = .008$, Bonferroni-corrected. Also, there was a main effect of amount of variance difference, $F(3,60) = 10.22$, $MSE = 405.40$, $p < .001$, $\eta_p^2 = .34$. Post-hoc comparisons showed that that when the variance difference was 10 pixels ($M = 73.33$, $SD = 19.65$) between target and the probe sets, accuracy was highest compared to a difference of 4 ($M = 53.33$, $SD = 22.57$, $p = .001$, Bonferroni-corrected) or 6 pixels ($M = 52.38$, $SD = 23.37$, $p < .001$, Bonferroni-corrected) but not different from the difference of 8 pixels ($M = 64.29$, $SD = 23.01$), all *ps* < 1.00, Bonferroni-corrected. Moreover, accuracy was higher in 6 pixels than 4 pixels (*p* = .037, Bonferroni-corrected) .Finally, interaction between two variables did not reach significance, *F*(3,60) = 1.27, *MSE* = 387.94, *p* = .29, η_p^2 = .06 (see Figure 39).

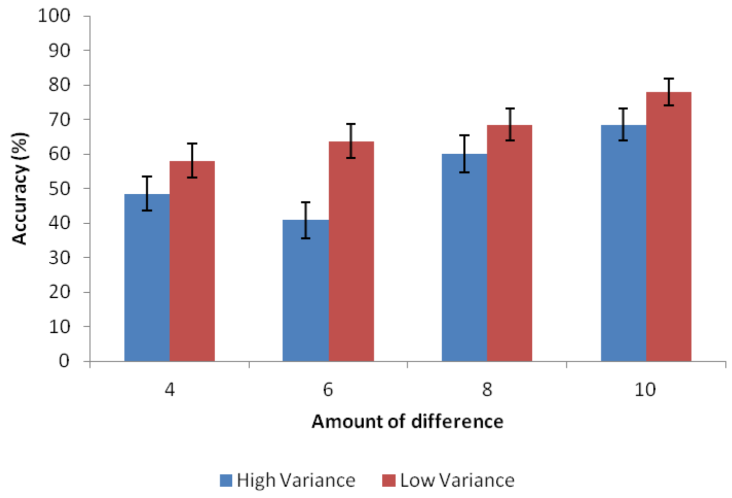


Figure 39. Percentage accuracy in each amount of variance in different conditions in Experiment 4A.

The RT pattern was consistent with the observed accuracy. There was a main effect of condition, *F*(1,18) = 6.52, *MSE* = 77662.73, *p* = .02, η_p^2 = .27. Pos-hoc comparisons showed that participants were slower in the high variance condition (*M* = 1399.78, *SD* = 187.734) compared to the low variance condition (*M* = 1298.30, *SD* = 241.07, *p* = .02, Bonferroni-corrected. Also, there was a marginally significant main effect of amount of variance difference, *F*(3,54) = 2.66, *MSE* = 55734.58, *p* = .058, η_p^2 = .13. Post-hoc comparisons showed that when the variance difference was 10 pixels (*M* = 1265.47, *SD* = 233.35) between target and the probe sets, RT was fastest compared to a difference of 4 (*M* = 1407.29, *SD* = 333.59, *p* = .049, Bonferroni-corrected), but not different from the difference of 6 pixels (*M* = 1368.75, *SD* = 298.72) or 8 pixels (*M* = 1382.47, *SD*= 323.51), all *ps* < 1.00, Bonferroni-corrected. There was no other significant difference. Finally, interaction between two variables did not reach significance, $F(3,54) = 1.26$, $MSE = 70031.57$, $p = .30$, $\eta_p^2 = .07$.[7]

Experiment 4B

In Experiment 4A, the same conditions of the variance estimation task were consisted of sets with the same mean and variance as the target set but included different individual circles. When we compared the range of the diameters of the individual circles within these pairs, we realized that 31 of 40 sets in the same condition came from out of the range of the target set. The number of outliers in each set was between 1 and 3. We speculated that participants might easily and quickly process these outliers and this might increase the miss rate. In order to eliminate this possibility, in Experiment 4B, in the same condition, we presented the target display with one change; individual circles making up the display were randomly displaced in the test phase and were presented in a different configuration. We expected this manipulation might decrease the tendency to say "different" in the same trials hence increase the accuracy.

Method

Participants

Thirty one students taking the introductory psychology course at Boğaziçi University

[7] When we omitted one participant whose overall accuracy was below chance level, the pattern of results remained the same.

participated in the study. The participants had normal or corrected-to-normal vision and gave written informed consent. All received course credit for their participation. Only the participants who performed 80% and above in the practice trials were included in the analysis. There were 21 participants who reached this criterion.

Results

Accuracy and Reaction Time Analysis

In order to examine whether performance was different in each condition, we compared percentage of accuracy in each condition. There was an effect of condition, $F(2,40) = 4.84$, $MSE = 180.90$, $p = .013$, $\eta_p^2 = .20$. Post-hoc comparisons showed that accuracy was lower in the high variance condition ($M = 51.67$, $SD = 17.20$) compared to the low variance condition ($M = 63.81$, $SD = 20.67$, $p = .047$, Bonferroni-corrected), but not different from same condition ($M = 61.55$, $SD = 15.38$), all $ps < 1.00$, Bonferroni-corrected. There was no other significant difference between conditions (see Figure 40).

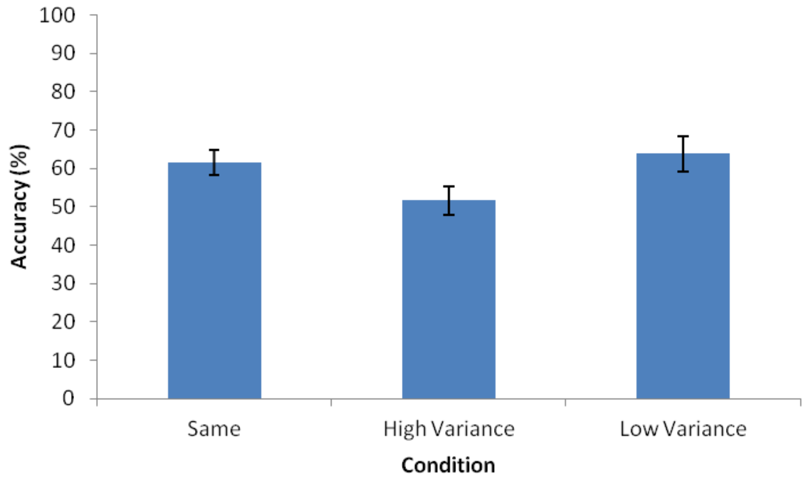


Figure 40. Percentage accuracy in each condition in Experiment 4B.

RT analysis revealed no main effect of condition, $F(2,40) = 2.41$, $MSE = 17550.82$, $p = .10$, $\eta_p^2 = .11$.

Amount of Difference in Terms of Variance

In order to examine whether as the variance difference between the target and probe sets increased, performance improved we conducted a 4 (Amount of Variance Difference (2, 4, 6, 8) X 2 (Condition: High vs. Low) within subjects ANOVA was conducted. There was a main effect of condition, $F(1,20) = 6.98$, $MSE = 887.86$, $p = .016$, $\eta_p^2 = .26$. Post-hoc comparisons showed that accuracy was lower in the high variance condition ($M = 51.67$, $SD = 17.20$) compared to the low variance condition ($M = 63.81$, $SD = 20.67$), $p = .016$, Bonferroni-corrected. Also, there was a main effect of amount of variance difference, $F(3,60) = 5.26$, $MSE = 512.78$, $p = .003$, $\eta_p^2 = .21$. Post-hoc comparisons showed that that when the variance difference was 10 pixels ($M = 68.10$, $SD = 28.37$) between target and the probe sets, accuracy was highest compared to a difference of 4 ($M = 50.95$, $SD = 26.36$, $p = .016$, Bonferroni-corrected) or 8 pixels ($M = 51.91$, $SD = 25.09$, $p = .021$, Bonferroni-corrected), but not different from the difference of 6 pixels ($M = 60.00$, $SD = 25.00$), all ps < 1.00, Bonferroni-corrected. Finally, interaction between two variables did not reach significance, $F(3,60) = .43$, $MSE = 363.10$, $p = .74$, $\eta_p^2 = .02$ (see Figure 41).

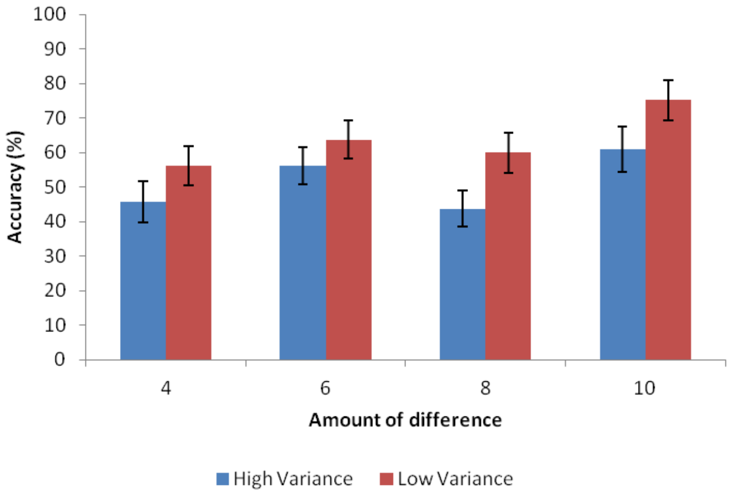


Figure 41. Percentage accuracy in each amount of variance in different conditions in Experiment 4B.

However, RT analysis revealed no significant results. Again for the purpose of brevity we did not report the results.[8]

Discussion

In Experiments 3 and 4 we investigated whether people can represent variance information accurately. We directly tested whether people were able to accurately estimate variance of a given set of circles. We speculated that as the variance difference between the distractor and the target sets increases, discrimination ability should improve. Moreover, reflecting on previous findings on the "downward attenuation" of the representation of variance, we hypothesized that accuracy should be higher when the distractor set had higher than lower variance than the target set. Contrary to our expectations, we found that performance dropped in high variance conditions. Moreover, we demonstrated that the amount of variance difference

[8]When we omitted 4 participants whose overall accuracy was below chance level, the pattern of results remained the same.

impacted performance, consistent with our expectations. We will elaborate on these in turn.

Interpretation of the Findings in Experiment 3

Contrary to our expectations we found no impact of our variance manipulation on accuracy. Our critical finding was that the amount of variance difference impacted performance, consistent with our expectations.

We speculated that the discrimination ability should improve as the variance difference between the distractor and the target sets increased. In Experiment 3A and B, the results revealed the amount of variance difference between target and distractors had an impact on the performance. Specifically, when the variance difference between the target and distractor was minimal (2 pixels), accuracy was lowest, partially supporting our predictions. Therefore, the results suggest that performance increased after the at least a 2 pixels difference was present. Since the accuracy levels are near chance level, we argue that participants were not able to discriminate when the variance difference between target and distractor was 2 pixels.

Based on these results, we were not able to reach a definitive conclusion. As mentioned above, we speculated that these null results may partially result from our task. In that task, we had presented participants first with a target set which was then followed by two sets, one of which was the same as the target set while the other was a different set with either higher or lower variance. The notion of "downward attenuation" suggests that the target set would be represented with a lower variance than it actually has. However, it is possible that we washed out this effect by presenting the target again at the test. Therefore, in Experiment 4, we used a different task.

Interpretation of the Findings in Experiment 4

To test the notion of "downward attenuation", in Experiment 4, we used a same/different task in which we first presented the target and then after a delay we presented the test stimuli for comparison. Also, we reduced exposure time to force participants to rely more on ensemble-type representations. In both experiments, we demonstrated that when the test display had higher variance than the target, performance was significantly lower than when it was identical to the target or lower; the difference between the conditions increased when we kept the individual items as well as the variance same in the same condition in Experiment 4B.

Our findings regarding the impact of the variance difference was similar across all experiments. As the variance difference between study and test displays increased performance was impacted. We also tested whether the discrimination ability should improve as the variance difference between the distractor and the target sets increased. Overall results suggest that the performance was highest when the difference between target and the probe set was 10 pixels, supporting our expectation.

We found no significant differences between the results of Experiment 4A (60% accuracy) and B (59% accuracy). The similarity of the results between Experiment 4A and B suggest that although the individual items within the sets in the same conditions were different, participants performance was similar. Therefore, it may be plausible to suggest that participants' responses were not merely affected by the individual items. The only difference between variance estimations tasks in two experiments was that in Experiment 4A, the target-probe pairs in the same condition had similar parameters in terms of variance and the mean size but they were consisted of different individual items while they were exactly the same in Experiment 4B. When compared the two experiments in terms of the accuracy levels in conditions.[9]

In any of the experiments, contrary to our expectations, we did not find any evidence for downward attenuation of variance. The accuracy level was lowest in the high variance condition (with a tendency to be around the chance level) among all conditions. In order to further examine the details of the sets we used in this condition, we looked at the range of the diameters of the individual circles within each condition (see Table 10). In the high variance condition, for each trial, the test stimuli included items that were outside of the initial range of the items constituting the initial study displays. These "outliers" at test could have led to a high proportion of "no" responses. However, if this had been the case, then performance would have been expected to be much higher in the high variance condition. In a similar vein, in the low variance condition there were only 5 sets with outliers, and performance was higher. The number of outliers ranged from 1 to 5. In contrast, only 5 of 20 sets in the low variance condition included outliers and in most conditions, there was only a single outlier. However, this finding works against our results. That is, if there is an effect of the number of outliers within a set on the performance, then the performance should be higher in the high variance condition compared to the low

[9]This relationship was tested by a 3 X 2 mixed design ANOVA. There was no main effect of experiment, $F(1,40) = 9.69$, $MSE = 376.56$, $p = .72$, $\eta_p^2 = .003$, and no interaction effects between condition and the experiment, $F(2,80) = 9.69$, $MSE = 167.76$, $p = .54$, $\eta_p^2 = .015$. In the same conditions, the accuracy levels in the same conditions were around 59% and 61% for Experiment 4A and B respectively.

variance condition than it actually is because outliers would increase the tendency to say “no”. However, results showed that performance was lower in the high variance condition compared to low variance condition.[10]

Table 10. The Number of the Sets with Outliers and the Total Number of Outliers in Each Condition

Condition	Outlier Yes	Outlier No	# of Outliers
Same	31	9	44
Different (HighVariance)	20	0	63
Different (Low Variance)	5	15	5
Amount of Difference			
4	9	1	14
6	6	4	15
8	5	5	18
10	5	5	21

Is There an Overestimation in Variance Representations?

Our findings suggest that performance dropped in the high variance condition. That is, it was hard to discriminate between target and probe sets when the probe set had higher variance than the target set. In Kareev et al.’s (2002) study which used a similar methodology to that of our Experiment 3, the results showed the opposite pattern. First they presented participants with a set of items followed by two sets: one

[10]Also, in order to see the effects of the number of outliers, we compared to performance on the sets with high number of outliers (4 or 5) and sets with low number of outliers (1, 2 or 3) in each condition. Both in Experiment 4A and B, although results were not significant, they revealed a general tendency of higher accuracy levels in the sets with high number of outliers (accuracy levels around 60% and 57%, respectively) compared to sets with low number of outliers (accuracy levels around 57% and 52%, respectively).

had the same variability with the target set while the other had either less or more variability compared to the target set. Similarly, the task was to judge which one of the probe sets was more similar to the target set. Results showed that the identical group was chosen more when it was given with a larger variable group (.66 of the cases) than with a smaller one (.49 of the cases). Therefore, they claimed that people estimate population variability smaller than it really is. That is, individuals are prone to consider the world to be regular or less variable.

One of the major differences between the present experiment and Kareev et al.'s (2002) study was the nature of the stimuli. They used sets of *3D* plastics disks and paper cylinders while we used *2D* drawings of circles. Therefore, it is possible that different mechanisms might be at role in representation of these different stimuli. Nevertheless, the accuracy rates were similarly low both in our and Kareev et al.'s experiments, suggesting that representation of variance does not seem to be an easy task.

Overall, our findings suggest that variance might be represented upwards rather than downwards in our specific context. That is, variance of a given set represented higher than it is. However, further research with more controlled stimuli and task is needed in order to draw a definitive conclusion.

Concluding Remarks

In the first set of experiments on mean size estimation, we reported people were highly accurate in estimating the mean size of a 9 same-sized circle as well as the size of a single circle. More importantly, we demonstrated that higher within-set variance resulted in poorer performance and an overestimation tendency. This pattern may be due to the fact that people utilized both subsampling strategies by focusing on specific items (such as extremes) and ensemble statistics. We speculated that the overestimation in higher variance may be due to a bias towards representing large circles larger than their actual size. This bias automatically leads to an overestimation in the representation of the variance itself.

When we tested the bias in variance estimations themselves more directly in the second set of experiments, we found a tendency to represent variance higher than it actually was, suggesting that variance might be represented upwards rather than downwards in our specific context. Therefore, when we combined the results of our study, we concluded that an overestimation of variance representations seems to be strong factor behind the overestimation in mean size estimations.

Appendix A

Informed Consent From

Bilgilendirilmiş Olur Formu

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi, Psikoloji Bölümü
Araştırmanın adı: Benzer Nesnelerin Grup Olarak Temsili
Araştırmacıların adı: Doç.Dr. Ayşecan Boduroğlu, Yelda Semizer
Adres: Boğaziçi Üniversitesi, Psikoloji Bölümü, 34342 Bebek-İstanbul
E-posta: aysecan.boduroglu@boun.edu.tr, yelda.semizer@boun.edu.tr

Sayın Katılımcı,

Bu deney bireylerin görsel temsil becerilerini araştırmaktadır. Sizden, bilgisayarda sunulacak olan bazı görsel veriler üzerinde çalışmanız ve kağıt üzerinde sunulan bazı sorulara cevap vermeniz istenmektedir. Deney, ortalama 1 saat sürecektir. Deneye katılmanız karşılığında size PSY 101 ya da PSY 241/242 dersinden 1 kredi verilecektir.

Bu araştırma bilimsel bir amaçla yapılmaktadır, toplanan veriler yayın amaçlı kullanılacaktır ve katılımcı bilgilerinin gizliliği esas tutulmaktadır. Adınız ve performansınız hiçbir şekilde eşleştirilmeyecektir.

Bu araştırmaya katılmak tamamen isteğe bağlıdır. Katıldığınız takdirde çalışmanın herhangi bir aşamasında sebep göstermeden çalışmadan ayrılma hakkına sahipsiniz, bu durumda kredi alma hakkınızı kaybetmeyeceksiniz. Araştırmayla ilgili sorularınızı deneyin sonunda bize yöneltebilirsiniz.

Bu önemli çalışmada bize yardımcı olmak isterseniz, lütfen aşağıdaki "İzin Formu'nu doldurup imzalayınız. Eğer 18 yaşından küçük iseniz, lütfen bu formu velinize imzalatıp araştırmacıya teslim ediniz. Çalışma hakkındaki bilgilendirmeyi okudum ve anladım.

Çalışmaya katılmak istiyorum ☐/ istemiyorum ☐

Velisi veya vasinin adı, soyadı ve imzası: ………………………………………………. (18 yaşından küçük katılımcılar için)

Bilgilendirilmiş Olur Formu'nun bir örneği tarafıma verildi.

Adı Soyadı:..

İmzası:...

Adresi:..

Telefonu: ....

E-posta: ....

Tarih (gün/ay/yıl):...../......./............

Appendix B

Highly Sensitive Person Scale

Duyusal Süreçler Hassasiyeti Anketi

ANKET YÖNERGESİ:

Lütfen alttaki ölçeği kullanarak, aşağıdaki her bir soruya kişisel olarak nasıl hissettiğinize dair 1 ile 7 arasında bir puan veriniz.

1	2	3	4	5	6	7
Hiç			Orta			Fazlasıyla

DUYUSAL SÜREÇ HASSASİYETİ SORULARI	PUAN
Güçlü duyusal uyaranlardan kolayca bunalır mısınız?	
Çevrenizdeki ince detayların farkına vardığınızı düşünüyor musunuz?	
Diğer insanların ruh hali sizi etkiler mi?	
Acıya karşı hassasiyet göstermeye eğilimli olduğunuzu düşünüyor musunuz?	
Kendinizi yoğun günlerde yatağa, karanlık bir odaya ya da uyaranlardan uzaklaşıp rahatlayabileceğiniz ve mahremiyete sahip olabileceğiniz bir yere çekilme ihtiyacı içinde buluyor musunuz?	
Kafeinin etkilerine karşı bilhassa hassas mısınız?	
Parlak ışık, güçlü kokular, kaba kumaşlar ya da yaklaşan siren sesi gibi şeylerden kolaylıkla bunalır mısınız?	
İçsel dünyanız zengin ve karmaşık mı?	
Yüksek sesten rahatsız olur musunuz?	
Sanat veya müzik sizi yoğun bir şekilde etkiler mi?	
Bazen sinir sisteminizin çok yıprandığı ve bu yüzden kendinizi sakinleştirmek zorunda kaldığınız oluyor mu?	
Özenli misiniz?	
Kolayca irkilir misiniz?	
Kısa bir zaman içinde yapacak çok işiniz varsa eliniz ayağınıza dolanır mı (dizleriniz titrer mi)?	
İnsanlar fiziksel bir ortamda rahatsız olduklarında, ortamı daha rahat hale getirmek için nelerin yapılması gerektiğini bilmeye eğilimli misiniz (ışığı veya oturulacak yeri değiştirmek gibi)?	

1 2 3 4 5 6 7

Hiç Orta Fazlasıyla

	PUAN
İnsanlar size aynı anda çok fazla şey yaptırmayı denediklerinde sinirlenir misiniz?	
Yanlış yapmaktan veya bir şeyleri unutmaktan kaçınmaya gayret eder misiniz?	
Şiddet içerikli film veya televizyon programlarından kaçınmaya özen gösterir misiniz?	
Etrafınızda süregelen çok fazla şey varsa hoş olmayan biçimde uyarılmış olur musunuz?	
Çok acıkmış olmak konsantrasyonunuzu ve ruh halinizi olumsuz yönde etkileyerek sizde güçlü bir tepkiye yol açar mı?	
Yaşamınızdaki değişiklikler sizi sarsar mı?	
Hassas veya güzel kokuları, tatları, sesleri, sanat ürünlerini fark eder ve onlardan zevk alır mısınız?	
Aynı anda süregelen pek çok şey olmasını tatsız (nahoş) bulur musunuz?	
Hayatınızı üzücü veya bunaltıcı durumlardan sakınacak şekilde düzenlemeye öncelik verir misiniz?	
Gürültü ya da kaotik görüntüler gibi yoğun uyaranlardan rahatsız olur musunuz?	
Bir görevi yürütürken rekabet etmek ya da izlenmek zorunda olduğunuzda, o görevde normalden daha kötü performans göstermenize yol açacak kadar gergin ya da güçsüz (beceriksiz, sarsak, ürkek) bir hale gelir misiniz?	
Çocukken ebeveynleriniz ya da öğretmenleriniz sizi hassas veya çekingen olarak görme eğiliminde miydiler?	

Appendix C

Spatial/Object Imagery Questionnaire

Görsel/Uzamsal İmgelem Anketi

ANKET YÖNERGESİ:

Lütfen alttaki ölçeği kullanarak, aşağıdaki her bir soruya kişisel olarak nasıl hissettiğinize dair 1 ile 5 arasında bir puan veriniz.

1	2	3	4	5
Hiç katılmıyorum				Tamamen katılıyorum

	PUAN
Üç boyutlu geometri konularında öğrenciliğimde çok iyiydim.	
Eğer mühendislik mesleği ve görsel sanatlar arasında seçim yapmam istenseydi, mühendisliği seçerdim.	
Mimari, resimden daha çok ilgilimi çeker.	
Zihnimdeki görüntüler (imgeler) çok renkli ve canlıdır.	
Bir ders kitabı okurken, renkli ve resimli illüstrasyonlar/örneklemeler yerine şematik diyagramları/grafikleri ve çizimleri tercih ederim.	
Zihnimdeki görüntüler (imgeler), detaylı resimlerden daha çok şematik temsillere benzer.	
Genellikle roman okurken tasvir edilen manzara/dekor ya da odanın net ve detaylı bir zihinsel resmini oluştururum.	
Fotografik bir belleğe sahibim.	
Üç boyutlu geometrik şekilleri rahat bir şekilde hayal edebilir ve zihnimde döndürebilirim.	
Belirli bir şey almak için tanıdık bir dükkana girdiğim zaman, alacağım şeyin tam yerini, üzerinde bulunduğu rafı, nasıl düzenlendiğini ve çevresindeki maddeleri kolaylıkla zihnimde canlandırabilirim.	
Normalde zihnimde spontane (birden) çok fazla görüntü oluşmaz; matematikteki gibi bazı problemleri çözmeye kalkıştığımda zihinsel görüntülerimi (imgelerimi) kullanırım.	
Zihnimdeki görüntüler (imgeler) çok canlı ve fotografiktir.	
Aşina olduğum bir binanın planını kolayca çizebilirim.	
Çok iyi bir tetris oyuncusuyum.	
Eğer okumak için mimarlık ve görsel sanatlar arasında bir seçim yapmam istenseydi görsel sanatları seçerdim.	

1	2	3	4	5
Hiç katılmıyorum				Tamamen katılıyorum

	PUAN
Farklı nesnelerin zihnimdeki görüntüleri, boyut, şekil ve renk açısından gerçekte gördüğüm nesnelere çok benzer.	
Bir arkadaşımın yüzünü hayal ettiğim zaman, zihnimde tamamen net ve canlı bir görüntü oluştururum.	
Teknik çizim yeteneklerim mükemmeldir.	
Başka bir kimsenin asla fark etmeyebileceği derecede çok fazla görsel detayı kolayca hatırlayabilirim. Örneğin, birisinin giydiği gömleğin ya da ayakkabının rengi gibi şeyleri kendiliğinden kavrayabilirim.	
Lisedeyken, sanata kıyasla geometride daha az zorlanırdım.	
Modern sanatlardaki resimler gibi canlı renklerden ve şekillerden oluşan resimlerden hoşlanırım.	
Bazen zihnimdeki görüntüler (imgeler) o kadar canlı ve sürekli oluyor ki onları görmezden gelmek zor oluyor.	
Soyut bir kavram hakkında (örneğin, 'bir bina') düşünürken, zihnimde somut belirli bir bina yerine soyut şematik bir bina veya binanın planını hayal edebilirim.	
Zihnimdeki görüntüler renkli ve resimli olmaktan çok şematiktir.	
Deneyimlediğim bir manzarayı gözlerimi kapatıp rahatça hayal edebilirim/resmedebilirim.	
Her şeyi görsel olarak hatırlarım. Bir akşam yemeğinde insanların ne söylediklerine kıyasla ne giydiğini anlatabilirim, oturma düzenleri ve nasıl göründükleri hakkında daha detaylı konuşabilirim.	
Üç boyutlu bir şeklin döndürüldüğünde tam olarak nasıl görüneceğini hayal etmekte zorlanırım.	
Zihnimdeki görsel temsiller (imgeler) sürekli olarak aklımdadır. Öylece oradadırlar.	
Grafik becerilerim mimaride kariyer yapmamı oldukça kolaylaştırır.	
Gerçekte hiç görmediğim bir radyo sunucusu veya DJ'i duyduğum zaman, kendimi genellikle o kişinin nasıl göründüğünü hayal ederken bulurum.	

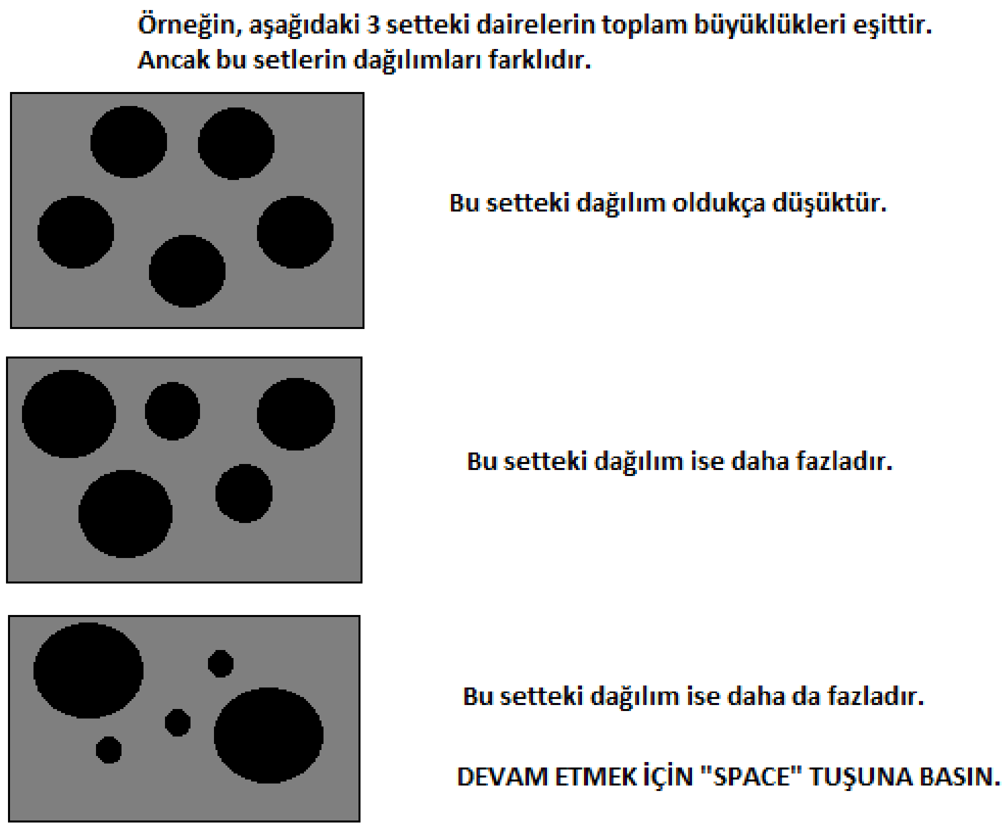
Appendix D

Instructions in the Variance Estimation Task (Experiment 3C)

Varyans Tahmini Görevi Yönergeleri (Deney 3C)

Bu çalışmada sizden ekranda beliren daire setlerini "dağılımları" açısından kıyaslamanız istenecektir.

Dairelerin dağılımı ile kastettiğimiz şudur: Dağılım bir setteki elemanların (örneğin dairelerin) büyüklüklerinin, ortalama değerden ne kadar farklı olduğuna verilen isimdir.



Önce ekranda beliren "+" işaretine odaklanınız. Bu işareti her gördüğünüzde işaret parmaklarınızı klavyede belirtilen tuşların üzerine ("D" ve "K" tuşları) koyunuz.

"+" işaretinin ardından ekranda bir daire seti belirecek. Bu daire setine odaklanınız. Bu set kaybolduktan sonra ekranın sağında ve solunda birer daire seti belirecek.

Bu setlerden birindeki dağılım bir önceki gördüğünüz setteki dağılım ile aynı iken, diğerindeki ise farklıdır. Bu üç setteki dairelerin toplam büyüklükleri ise eşittir.

Sizden istediğimiz bir önceki gördüğünüz setteki ile aynı dağılıma sahip olan setin hangi tarafta olduğuna karar vermenizdir.

Eğer soldaki ise "D" tuşuna basınız. Eğer sağdaki ise "K" tuşuna basınız.

Setler bir süre sonra ekrandan kaybolacaktır. Tuşa isterseniz iki set ekranda iken, isterseniz setler ekrandan kaybolduktan sonra basabilirsiniz.

Kıyaslama yaparken dairelerin yerlerine dikkat etmenize gerek yoktur.
Önemli olan dairelerin benzerliğidir.

Önce biraz alıştırma yapalım. Her deneme sonrası cevabınızın doğruluğu ile ilgili bilgilendirileceksiniz.

Eğer soldaki ise "D" tuşuna basınız. Eğer sağdaki ise "K" tuşuna basınız.

>>>>>Alıştırmalar<<<<<<

Alıştırmaları bitirdiniz. Şimdi gerçek denemelere başlayabilirsiniz. Artık cevaplarınız doğruları hakkında bilgi verilmeyecek, denemeler birbiri arkasına hızlı bir şekilde gelecektir.

Deney sırasında bir kere dinlenme hakkınız olacak ve bu size bildirilecektir. Sorunuz varsa lütfen deney görevlisini çağırınız.

Appendix E

Instructions in the Variance Estimation Task (Experiment 4A)

Varyans Tahmini Görevi Yönergeleri (Deney 4A)

Birazdan ekranda dairelerden oluşan setler göreceksiniz.
Bu setler birbiri ile kıyaslandığında her birinin farklı bir dağılımı olduğu görülebilir.
Bu çalışmada sizden ekran göreceğiniz daire setlerini "dağılımları" açısından kıyaslamanız istenecektir.

Size öncelikle dağılımın ne anlama geldiğini anlatalım:

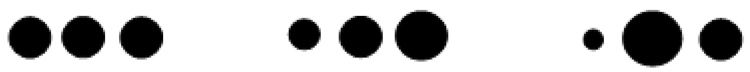
Dağılım, bir setin elemanlarının birbirlerinden ne kadar farklı olduklarına verilen isimdir. Örneğin, aşağıdaki 3 seti ele alalım:

Set A = {5, 5, 5}   Set B = {4, 5, 6}   Set C = {1, 5, 9}

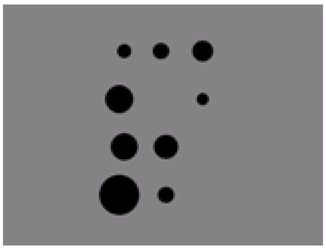
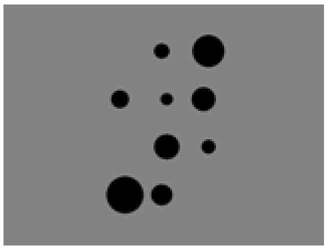
Her bir sette eşit sayıda elman vardır. Bu setlerdeki sayıların ortalaması 5 olmasına rağmen, dağılımları farklıdır. A setindeki sayılar birbirinin aynıdır bu sebeple en küçük dağılım A setindedir. B setindeki sayılar birbirlerinden biraz farklıdır bu sebeple B setindeki dağılım A setine kıyasla daha fazladır. C setindeki sayılar birbirlerinden oldukça farklıdır bu sebeple C setindeki dağılım diğer iki sete kıyasla en fazladır.

Şimdi de dağılımı daireler üzerinden inceleyelim:

Set A: Set B: Set C:



Her bir sette eşit sayıda elman vardır. Bu setlerdeki dairelerin ortalama büyüklükleri eşit olmasına rağmen, dağılımları farklıdır. A setindeki daireler birbirinin aynıdır bu sebeple en küçük dağılım A setindedir. B setindeki daireler birbirlerinden biraz farklıdır bu sebeple B setindeki dağılım A setine kıyasla daha fazladır. C setindeki daireler birbirlerinden oldukça farklıdır bu sebeple C setindeki dağılım diğer iki sete kıyasla en fazladır.

Yukarıda gördüğünüz 2 set birbirlerinden farklı büyüklükte elemanlara sahip olmalarına rağmen, set A ve B aynı dağılıma sahiptir.

.

Buraya kadar okuduklarınız hakkında bir sorunuz var mı? Varsa lütfen deney görevlisine sorunuz.

Şimdi ise biraz alıştırma yapalım. Birazdan ekranda yan yana iki daire seti göreceksiniz. Sizden istediğimiz bu daire setlerini dağılımları açısından kıyaslamanızdır. Hangi daire setinin dağılımı daha fazla ise onu işaretleyiniz. Sağdaki ise sağ ok tuşuna, soldaki ise sol ok tuşuna basınız. Şimdi iki elinizin işaret parmaklarını bu tuşlara yerleştiriniz.

Hazır olduğunuzda başlayınız.

>>>>>Alıştırmalar<<<<<<

Alıştırmaları tamamladınız. Sorunuz varsa lütfen deney görevlisini çağırınız.

Artık gerçek göreve başlayabilirsiniz. Bu görevde sizden istediğimiz ekranda göreceğiniz daire setlerini dağılımları açısından kıyaslamanızdır. Önce ekranda bir daire seti belirecek. Set kaybolduktan sonra onun yerinde yeni bir daire seti belirecek. Bu set bir önceki ekranda görmüş olduğunuz setten farklı bir set olacak. Ancak, sizden istediğimiz bu setin dağılımını göz önünde bulundurmanız.

Eğer bu setin dağılımı bir önceki gördüğünüz set ile aynı ise "k", farklı ise "d" tuşuna basınız.

Lütfen yanıtlarınızı olabildiğince hızlı ve doğru bir şekilde veriniz. Şimdi sol işaret parmağınızı "d" tuşuna, sağ işaret parmağınızı "k" tuşuna yerleştiriniz.

Hazır olduğunuzda başlayınız.

REFERENCES

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychological Science, 21*, 560–567.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122-131.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics, 70*(7), 1325-1326.

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology, 73,* 345-68.

Aron, E. N., Aron, A., & Davies, K. M. (2005). Adult shyness: The interaction of temperamental sensitivity and an adverse childhood environment. *Personality and Social Psychology Bulletin, 31*(2), 181-97.

Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance, 3*, 109-123.

Benham, G. (2006). The highly sensitive person: Stress and physical symptom reports. *Personality and Individual Differences, 40*, 1433-1440.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-Spatial Imagery: A new self-report imagery questionnaire. *Applied Cognitive Psycholgy, 20,* 239-263.

Boduroglu, A., Ng, A., Mueller, S. & Shah, P. (in preparation). Representation resolution is correlated with visual and spatial working memory capacity (draft available).

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384-392.

Charras, P., Brod, G., & Lupiáñez, J. (2012). Is 26 + 26 smaller than 24 + 28? Estimating the approximate magnitude of repeated versus different numbers. *Attention, Perception, & Psychophysics, 74,* 163-173.

Charras, P., & Lupiáñez, J. (2009). The relevance of symmetry in line length perception. *Perception, 38,* 1428–1438.

Chong, S. C., Joo, S. J., Emmanouil, T., & Treisman, A. (2008). Statistical processing: not so implausible after all. *Perception & Psychophysics, 70,* 1327–1334.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67,* 1–13.

Chong, S.C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45,* 891–900.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-114.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37,* 3181-3192.

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception and Psychophysics, 70,* 789-794.

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across future dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70*(6), 946-954.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

Evans, J. St. B. T., & Pollard, P. (1985). Intuitive statistical inferences about normally distributed data. *Acta Psychologica, 60,* 57-71.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,* 21–38.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464-472.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*, 751-753.

Jagiellowicz, J. et al. (2010). The trait of sensory processing sensitivity and neural responses to changes in visual scenes. *Social Cognitive Affecitive Neuroscience, 6*, 38-47.

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 131*(2), 287-297.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*, 278 –287.

Kimchi, R. (1992). Primacy of wholistic processing and global/local: A critical review. *Psychological Bulletin, 112*, 24-38.

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition, 33*, 710-726.

Levin, I. P. (1975). Information integration in numerical judgments and decision processes. *Journal of Experimental Psychology-General, 104*, 39-53.

Lovie, P., & Lovie, A. D. (1976). Teaching intuitive statistics I: Estimating means and variances. *International Journal of Mathematical Education in Science and Technology, 7*(1), 29-39.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Marchant, A. P., & De Fockert, J. W. (2009). Priming by the mean representation of a set. *The Quarterly Journal of Experimental Psychology, 62*(10), 1889-1895.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17-24.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences, 9*(6), 296-305.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics, 69*(8), 1324–1333.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*, 772–788.

Oliva, A., & Schyns, P. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176-210.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23-36.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739-744.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29-46.

Pollard, P., 1984. Intuitive judgments of proportions, means and variances: A review. *Current Psychological Research and Reviews, 3*, 5-18.

Price, M. C. (2009). Spatial forms and mental imagery. *Cortex, 45,* 1229-1245.

Rosenbaum, D. A., & Kornblum, S. (1982). A priming method for investigating the selection of motor responses. *Acta Psychologica*, *51*(3), 223-243.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.) *Advances in experimental social psychology* (Vol. 10, pp. 173–220). San Diego, CA: Academic Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide* Pittsburgh: Psychology Software Tools, Inc.

Simons, D. J., & Myczek, K. (2008). Average size perception and allure of a new mechanism. *Perception & Psychophysics*, *70*(7), 1335-1336.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119*(4), 807-830.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505.

Van Opstal, F., de Lange, F. P., & Dehaene, S. (in press). Rapid parallel semantic processing of numbers without awareness. *Cognition, xxx*, xxx-xxx.