

OUTLIER AND ENSEMBLE REPRESENTATIONS OF VISUAL DISPLAYS

İREM YILDIRIM

BOĞAZIÇI UNIVERSITY

2016

OUTLIER AND ENSEMBLE REPRESENTATIONS OF VISUAL DISPLAYS

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

by

İrem Yıldırım

Boğaziçi University

2016

DECLARATION OF ORIGINALITY

I, İrem Yıldırım, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date27/05/2016

ABSTRACT

Outlier and Ensemble Representations of Visual Displays

Ensemble representations are argued to benefit efficient outlier processing. However, limited research has investigated the relationship between outlier and ensemble representations which was the goal of this thesis. Eight heterogeneously-sized circles were presented and viewers reported the mean size or the largest/smallest size across different blocks. The largest/smallest size could be the outlier or not in the particular distribution of a display. In a separate block that only had displays with outliers, participants reported either the mean or the largest/smallest size. Since the relevant task was indicated after display, in this mixed block, viewers were encouraged to concurrently focus on the ensemble along with the outlier. I specifically investigated how the presence of an outlier impacts the ensemble representations; whether outliers are better represented than non-outliers; whether people can form ensemble and outlier representations simultaneously. Results showed that an outlier was always excluded from mean estimations. Only a larger outlier decreased the mean size accuracy. Also, representations were more veridical when the largest/smallest item was an outlier than when it was not. Last, attending to the mean and outlier size concurrently did not yield a cost in the mean size and larger outlier precision suggesting that viewers were able to process outlier and mean size concurrently. However, when the outlier was smaller, the outlier estimates were less accurate. I conclude that an outlier is efficiently processed and somewhat downplayed in ensemble representations. Nevertheless, ensemble representations may not foster the processing of outliers that are perceptually harder to detect.

ÖZET

Görsel Ekranların Uç Nesne ve Özet Temsilleri

Özet temsillerin uç nesnelere etkin bir şekilde işlenmeye yardımcı olduğu önerilmektedir. Fakat, sınırlı sayıda araştırma uç nesne ve özet temsiller arasındaki ilişkiyi incelemiştir. Bu tezin amacı uç nesne ve özet temsiller arasındaki etkileşimi araştırmaktır. Katılımcılara farklı boyutlarda 8 daire gösterildi ve ortalama boyutu veya en büyük/küçük boyutu ayrı bloklarda bildirmeleri istendi. En büyük/küçük boyut, dairelerin dağılımına göre, ekranda uç nesne veya değildi. Sadece uç nesne bulunan setler gösterilen ayrı bir blokta, katılımcılar ya ortalamayı ya da en büyük/küçük boyutu belirttiler. Geçerli görev daire seti kaybolduktan sonra bildirildiği için bu karma blokta katılımcıların hem bütüne hem de uç nesneye aynı anda odaklanmaları desteklendi. Özel olarak; uç nesne bulunmasının özet temsilleri nasıl etkilediği, uç nesnelerin uç nesne olmayanlardan daha iyi temsil edilip edilmedikleri, insanların özet ve uç nesne temsillerini aynı anda oluşturup oluşturamadıkları incelendi. Sonuçlar gösterdi ki uç nesne olan boyut ortalama boyut tahminlerine dahil edilmedi. Yalnızca daha büyük uç nesne, ortalama boyut kesinliğini düşürdü. Ayrıca, en büyük/küçük daire temsilleri uç nesne olduğunda, olmadığı duruma göre, daha gerçeğe yakındı. Son olarak, bütüne ve uç nesneye aynı anda dikket vermek ortalama boyut ve daha büyük uç nesne kesinliğinde bir bedele yol açmadı. Ancak, uç nesne daha küçük olduğunda, uç nesne tahminleri daha hatalıydı. Varılan sonuç, uç nesnelerin etkin biçimde işlendiği ve bir şekilde özet temsillerde önemsenmediğidir. Yine de, özet temsiller, algısal olarak tespiti zor olan uç nesnelerin işlenmesini teşvik etmiyor olabilir.

ACKNOWLEDGMENTS

I am thankful to The Scientific and Technological Research Council of Turkey (TÜBİTAK) for providing scholarship throughout my MA education. Specifically, I was financially supported under 2211 – Domestic Graduate Scholarship Program and 1001 – Scientific and Technological Research Projects Funding Program. The latter project (113K217) was a grant awarded to Dr. Ayşecan Boduroğlu.

I would like to express my deepest gratitude to my advisor, Dr. Ayşecan Boduroğlu, for her encouragement, understanding, and expert guidance throughout my research. Her full support, patience and wisdom underpin the success here.

I would also like to thank Dr. İnci Ayhan and Dr. Burak Güçlü for having served on my thesis committee. I appreciate their helpful questions and comments on many occasions.

I am sincerely grateful to my fellow graduate students, Aysu Mutlutürk and Erdem Ozan Meral, who took their time to talk with me about research ideas. They were influential on my academic exploration. I also appreciate the assistance of Öykü Üner and Hande Sodacı in data collection.

Certainly, special thanks go to my parents, sister, and brother-in-law for their unconditional and continuous love. I always felt their sincere support which was the key motivation for pursuing and completing my MA degree.

DEDICATION

To my mother, Bahriye Yıldırım

For teaching me how to be enthusiastic

To my father, Üzeyir Yıldırım

For firing me up towards knowledge

To my sister, Didem Özhan

For encouraging me to believe in myself

To my niece to-be-born, Göksel Özhan

For inspiring me to live life to the fullest

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: BACKGROUND	3
CHAPTER 3: THE AIM, QUESTIONS, AND HYPOTHESES	6
CHAPTER 4: METHOD	9
4.1 Participants	9
4.2 Materials and stimuli	9
CHAPTER 5: RESULTS	14
5.1 The mean size estimation	14
5.2 The largest/smallest size estimation	16
5.3 The outlier and mean size estimation	18
5.4 Estimation biases as a control.....	19
CHAPTER 6: DISCUSSION.....	21
APPENDIX: INFORMED CONSENT FORM	27
REFERENCES.....	28

LIST OF FIGURES

Figure 1. An illustration of a mean size discrimination experiment.....	3
Figure 2. Graphs showing the awareness of the range of stimuli	5
Figure 3. An example trial (a) in the single task and (b) mixed blocks of Experiment 1A.....	11
Figure 4. Illustration of displays in Experiment 1B	13
Figure 5. Graphs depicting (a) accuracy and (b) bias towards local over global mean in the mean size reports.....	16
Figure 6. Graphs depicting (a) accuracy and (b) bias towards mean in the largest/smallest size reports.....	17
Figure 7. Graphs depicting accuracy for displays with an outlier.....	19

CHAPTER 1

INTRODUCTION

Imagine that you are camping in a rich natural environment. There is a variety of plants and animals. You decide to look around to enjoy being in nature. Would you look at every tree, every bush, every flower, or every animal? You would probably not. Instead, your eyes would scan the scenes as you move forward and turn your head around. You would notice the characteristics of the vegetation of that land for sure; whether it has steep trees with large bright green leaves and slender branches, short but broad bushes like saying “you shall not pass”, and warm colored flowers of every kind. You would see the insect colonies, and dun colored horses, and flying birds with gorgeously large wings. You would definitely be aware of that land with those features giving you the gist of the scene to experience those peaceful moments. What else would you notice as specific things? Can it be the sideling tree that looks against gravity or the tree covered by dry tiny yellowish leaves among ones with large and fresh leaves? You may realize the specific bush which is the smallest among others so that you can enter a path and go deeper into the forest. Perhaps, the ice blue flower that is unique with its color among yellow, red, and orange ones is the specific beauty you would find discernible. What about animals? Would you be able to scrutinize each of the insect colonies you come across or each dun horse with their own color and string patterns? Do you believe you would look through each bird’s wings seeming wonderful on its own? I suspect that the giant insect colony that might scare you or the white horse that might amaze you or the bird possessing the incredibly tiniest wings that might surprise you would catch your focus. All of

these noticeable creatures and plants have one thing in common; their visual features are very different from the other exemplars of their community living in that land.

The above story narrates the background idea of the current thesis. That is, when the visual system sees an ensemble of objects, it ends up with a summary of those objects at first hand which is the gist of a natural scene, average orientation of trees like being steep, color tendency of flowers such as warm colors, or average size of the colonies and so on. By doing that, visual system understands statistical regularities of those objects and represents the ensemble before than representing each exemplar in detail. Perceiving and encoding specific instances is influenced by that summary representation. One way that this influence show itself might be the relative involvement of the specific exemplar in the summary representation.

Particularly, the outlier item that is separable enough from the rest of the ensemble might be the best candidate to be attended and encoded. The outlier item adds so much variance to the ensemble and breaks the statistical regularity that all other items established. In the above story, the sideling tree among steep ones, or the ice blue flower among warm colored others, or the giant insect colony among medium colonies are the outlier items that were ruining the representations of their ensembles. The catchiness of them would probably disappear if the other trees would also be sideling, or if the other flowers would also have blue, purple, and turquoise bracts, or if the other colonies would be big. Thus, the statistical regularity of the ensemble, in this story, functions as to make those outliers more detectable. The essence of current thesis is to understand the outlier and ensemble processing of visual displays and their influences on each other.

CHAPTER 2

BACKGROUND

Studies have revealed that people are able to extract compressed information from a visual display (for reviews see Alvarez, 2011; Haberman & Whitney, 2012) such as the gist of a natural scene (Oliva & Torralba, 2006) or the average (e.g. Chong & Treisman, 2003) and variance (Morgan, Chubb, & Solomon, 2008; Peterson & Beach, 1967; Semizer, 2013) of simple features. These so-called ensemble (statistical summary) representations can be formed very rapidly (Chong & Treisman, 2003; Whiting & Oriet, 2011) and are, unlike individual item representations, very accurate (Ariely, 2001; Corbett & Oriet, 2011). People can extract robust ensemble representations with both low-level features (e.g. size: Albrecht & Scholl, 2010; Ariely, 2001; orientation: Attarha & Moore, 2015; Parkes, Lund, Angelucci, Solomon & Morgan, 2001; brightness: Bauer, 2009; location: Alvarez & Oliva, 2008; Mutlutürk & Boduroglu, 2016) and high-level stimuli (e.g. face identities: de Fockert & Wolfenstein, 2009; emotional expression and gender of faces: Haberman & Whitney, 2007; 2009; biological motion: Sweeny, Haroz, & Whitney, 2013). A schematic demonstration of a statistical summary task can be seen in Figure 1.

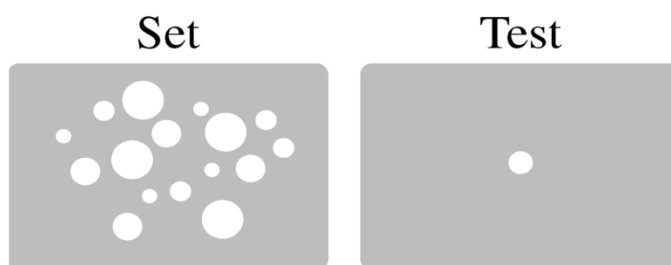


Figure 1. An illustration of a mean size discrimination experiment

Note: Adapted from “Seeing Sets: Representation by Statistical Properties”, by D. Ariely, 2001, *Psychological Science*, 12(2), p. 158. Copyright (2001) by American Psychological Society. The task was to indicate whether the mean size of the presented set was smaller or larger than the test item.

Current models of visual perception have suggested that such summary representations may serve multiple functions, such as form the basis of scaffolding in scene processing (Bar, 2004; Oliva & Torralba, 2007; Torralba & Oliva, 2003). It has also been hypothesized that these representations may foster the tagging and efficient processing of outliers (Alvarez, 2011; Haberman & Whitney, 2012). Since outliers, by definition, differ from a set that may typically consist of items with relatively similar values on a particular dimension, extraction of a statistical summary may consequently highlight the odd item. Rosenholtz and her colleagues argued that the successful identification of a target among distractors during visual search is dependent on the statistical summaries of the pooled distractors (Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Specifically, when the target was further away from the mean of the non-identical distractors, it was detected faster (Rosenholtz, 1999). Similarly, in one of the early studies, Ariely (2001) asked viewers to report whether a test item belonged to the studied set. Performance was above chance level when the test item was outside the range of items in the studied set (Figure 2). That was, viewers were able to correctly reject a test item if it fell outside the stimuli range. This suggested people were aware of the range of items in a given set; thus, an outlier beyond the range of a coherent ensemble can be identified.

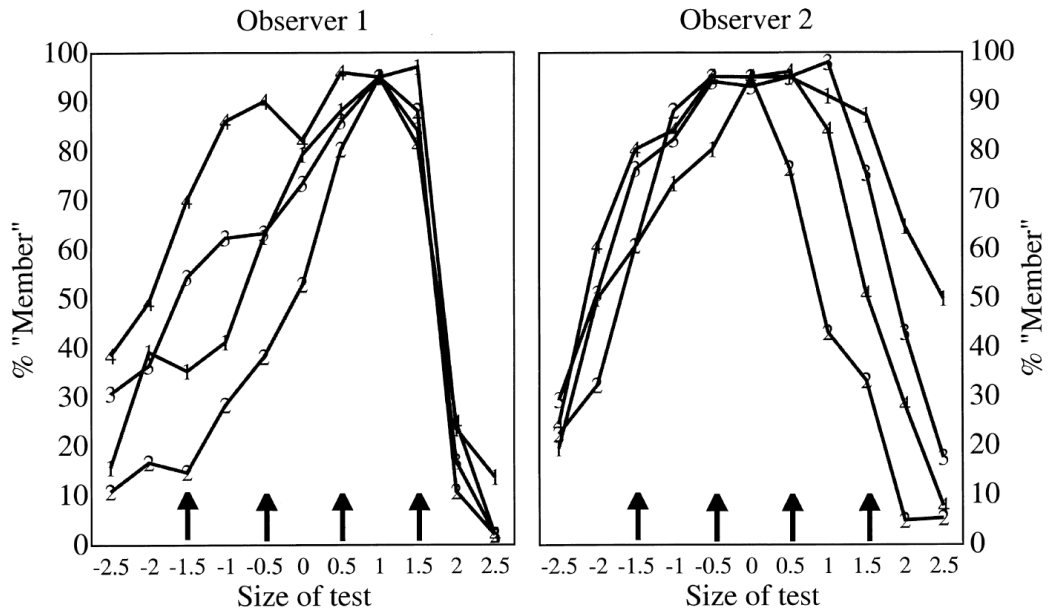


Figure 2. Graphs showing the awareness of the range of stimuli

Note: Adapted from "Seeing Sets: Representation by Statistical Properties", by D. Ariely, 2001, *Psychological Science*, 12(2), p. 158. Copyright (2001) by American Psychological Society. The task was to decide whether a single test item was a member of the presented set or not. The y-axis shows the percentage of member responses which could be correct or incorrect. The arrows on the x-axis indicate the members of the set. Performance is below 50% for the spots without arrows indicating observers' above chance performance in correctly rejecting stimuli beyond the range of stimulus set.

CHAPTER 3

THE AIM, QUESTIONS, AND HYPOTHESES

In light of these findings, in the present thesis I specifically investigated the interaction between outlier and ensemble representations. Across two experiments, I addressed three questions: 1) how does the presence of an outlier influence the statistical summary representation; 2) are outliers more precisely represented than non-outlier items; 3) can ensemble and outlier representations be formed concurrently?

In order to address these questions, I used a novel paradigm in which I presented participants heterogeneously sized circle displays and asked them to report the size of either the largest (Experiment 1A)/the smallest (Experiment 1B) item or the mean size. On half of the trials the largest/smallest item was an outlier; these outliers were defined statistically based on the particular size distribution in that display. Unlike many ensemble tasks that utilize a two-alternative forced choice task (e.g. Chong & Treisman, 2005), I allowed participants to adjust their estimates continuously, and was able to compare highly precise estimates across conditions. In both experiments, participants first did the item size estimation and mean size estimation in a counterbalanced yet blocked fashion. Then in a separate block, these trials were randomly intermixed to force viewers to attend both pieces of information simultaneously and help determine to what extent mean and outlier representations can be concurrently formed. In both experiments, I was interested in not only the precision of the representations, but also whether there were any systematic biases.

I expected mean size estimations to be more erroneous when displays contained an outlier. Earlier work in our lab had shown that viewers represented

mean size less precisely when there was greater variability in a circle set (Semizer, 2013). Since the presence of an outlier naturally increases variance, I expected to replicate this earlier work. The only other study focusing on sets with outliers used a mean facial emotion estimation task and showed that people discounted outliers (Haberman & Whitney, 2010). Thus, I also expected mean estimates to be biased towards the local mean (excluding the outlier) rather than the global mean (including the outlier).

The second question I was interested in was whether outlier representations were more precise than non-outlier representations. There is some earlier work to suggest that items that deviate from the gist of an ensemble may indeed be better represented. Brady & Tenenbaum (2010; 2013) proposed that viewers catch the regularity in a display and perceptually organize items accordingly. They argued that this allows outliers to also be encoded since they break the regularity and provide more information about a display together with the ensemble representation. Indeed, Brady and Tenenbaum (2010) demonstrated that changes in outlier items were detected more easily in a change detection paradigm. Regarding these findings, I predicted outliers to be more accurately represented than non-outlier ensemble members. I specifically compared representation precision across trials for items that were identical in size (largest or smallest) but were either an outlier or not based on the rest of the distribution. Furthermore, I also predicted that the non-outlier estimate may show more of a bias towards the global mean compared to the outlier item. Earlier researches from our lab (e.g. Mutlutürk & Boduroglu, 2016; Yildirim, Öğreden & Boduroglu, 2015) and others (e.g. Brady & Alvarez, 2011) have shown that people tend to be biased towards the summary statistics when they only have access to a relatively imprecise representation. In a somewhat related vein,

Huttenlocher, Hedges, and Vevea (2000) found that inexact item representations were pulled towards the center of a learned category's frequency distribution. They also demonstrated that representations of low probable members (somewhat like outliers) of that category were less biased towards the center.

My third question was related to whether outlier and mean size representations could be concurrently formed. It has been argued that ensemble representations may be automatically extracted (e.g. Chong, Joo, Emmanouil, & Treisman, 2008; but also see Marchant, Simons & de Fockert, 2013; Myzczek & Simons, 2008). Similarly, outliers (i.e. singletons) in visual search tasks have been shown to "pop out" among distractors; it has been argued that they are processed pre-attentively (Itti & Koch, 2000; 2001; Treisman, 1982; 1988; Wolfe, 1994). Bridging these two findings with the idea that ensemble representations may lead to perceptual organization and tagging of the outlier, I specifically looked at performance on the final mixed block. In the mixed block, in both experiments, all displays had an outlier and after the offset of the display participants were randomly cued to report either the mean size or the largest/smallest size (i.e. the outlier item). If these representations can be formed concurrently, then error levels for the single block trials and the mixed block trials should be comparable and there should be no additional cost associated with attending to both pieces of information. Specifically, null effect of concurrent focus should present in both mean size and outlier size estimations.

CHAPTER 4

METHOD

4.1 Participants

Twenty nine Bogazici University students (11 female, $M_{\text{age}} = 21.28$ years, age range: 18 – 30 years) were recruited via personal contact or in exchange for course credits. All participants reported normal or corrected-to-normal vision and all provided written informed consent (see Appendix). They participated to both Experiment 1A and 1B on separate days. The order of Experiment 1A and 1B was counterbalanced across participants. Three participants were identified as extremes on stem-and-leaf plots of a block and were removed from the analyses. Thus, the reported results reflect data from 26 participants (10 female).

4.2 Materials and stimuli

The experiment was presented via the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) on a 17" computer screen. Participants were seated approximately 57 cm away from 640×480 pixels screen. From this distance, 20 pixels corresponded to 1° of visual angle. Each experiment consisted of 3 blocks and participants were asked to either reproduce the size of the largest (Experiment 1A)/smallest (Experiment 1B) item or the mean size of a presented display. The first two blocks were “pure” blocks such that participants engaged in one of these tasks; task type was counterbalanced across participants. The third block was a mixed block; participants were randomly post-cued to report either the largest/smallest size or the mean size. Each block consisted of 16 practice trials with feedback and 112 experimental trials without feedback. In the single task blocks, participants estimated

the largest/smallest size and the mean size of Outlier+ and Outlier– displays. In the mixed block, all trials had an outlier.

4.2.1 Experiment 1A

In the single task blocks, half of the trials were circle displays with an outlier which was the largest among smaller circles and in the other half of the trials, there was the largest circle equivalent in size but was not an outlier. There were no more than 4 successive trials with either an outlier (Outlier+) or without an outlier (Outlier–). In the mixed block, all displays had an outlier and in half of the trials, I asked the largest size and in the other half I asked the mean size. There were no more than 4 consecutive trials asking for the same task.

At the beginning of each trial (see Figure 3 for an example trial), participants saw a fixation plus at the center of the screen for 500 ms. The fixation was followed by the study display presented for 250 ms. A response screen with a test circle immediately appeared after the study display. Participants adjusted the test circle placed at the center of the screen by pressing key (D to make the response circle diameter smaller and K to make it larger by 0.1°) to match it to the required size. The correct size was achieved by at least five and at most ten key presses. To indicate the relevant task in the mixed block, either the word “MEAN” or “LARGEST” was presented at the top of the screen in blue and red, respectively. A blank screen appeared during the 1000 ms inter-trial interval.

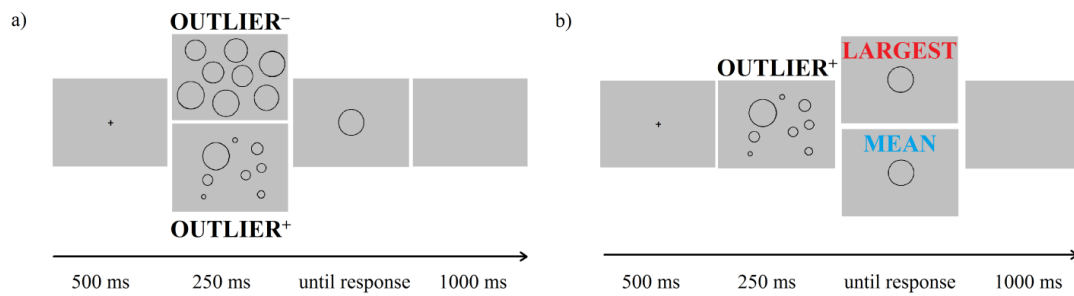


Figure 3. An example trial (a) in the single task and (b) mixed blocks of Experiment 1A

In each circle display, I presented eight black outlined circles on a gray background.¹ All of the circles were uniquely sized. On the basis of the psychological scale which is a power function with the 0.76 exponent (Teghtsoonian, 1965), the diameters of any two circles differed by at least 0.2° difference in order to be perceived as different in size.

In the Outlier+ displays circles were placed such that there was a 3° wide horizontal strip from the top and the bottom and a 4° wide vertical strip from the left and the right of the screen. All of the eight circles were positioned randomly on the rest of the screen with the constraint that inter-item distance was a minimum 1° . To ensure that the outlier was not spatially segregated from the remaining items, the distance of outlier to the two closest circles were the minimum two distances between items in the display. Diameters of the seven member circles, the items constituting the ensemble except for the outlier, were randomly selected between 0.8° and 3.2° of visual angle. The diameter of the one largest outlier circle was calculated by adding 5.6° to the local mean diameter, which was the average of the seven member diameters. This meant that the outlier circle's diameter was at least 2.25 standard deviations more than the global mean. This also meant that the difference between the diameters of local and global mean was always 0.7° . Possible

¹ Circles were outlined instead of filled in order to minimize luminance differences between Outlier+ and Outlier- displays.

global mean diameters were 2.3° , 2.5° , 2.7° , and 2.9° while possible outlier diameters were 7.2° , 7.4° , 7.6° , and 7.8° of visual angle. These displays with 4 unique mean and outlier sizes were equally often presented in each block. In the mixed block in which only Outlier⁺ displays were presented, no more than 4 consecutive trials had the same mean and outlier size displays.

In the Outlier⁻ displays, there was a 0.3° wide horizontal script from the top and the bottom and 0.4° wide vertical strip from the left and the right of the screen. By keeping a minimum 0.1° inter-item distance, seven circles were located randomly on the rest of the screen but the largest circle was located in the same area as in Outlier⁺ displays. The maximum two inter-item distances were not between a circle and the largest circle to ensure that the largest size was not be perceived as spatially segregated. The largest circle sizes of Outlier⁻ displays were matched with the possible largest sizes of Outlier⁺ displays. Hence, possible largest circle diameters were 7.2° , 7.4° , 7.6° , and 7.8° of visual angle. However, the largest size was not an outlier given the distribution of the other circles. Diameters of other seven circles were selected among 13 alternatives as in Outlier⁺ displays but decreasing by 0.2° from the largest size. That is, for the display with the largest size of 7.8° , the rest of the circles were randomly sized between diameters of 5.2° and 7.6° . Likewise, if the largest size was 7.2° , the other seven circles' diameters ranged from 4.6° to 7° . This meant that the largest size was at most 1.5 standard deviations more than the mean. Displays with the largest size of 7.2° and 7.4° had mean diameter of either 6.0° or 6.2° while displays with the largest size of 7.6° and 7.8° had mean diameter of either 6.4° or 6.6° . Thus, I had 4 possible means as in Outlier⁺ displays. These displays with 4 unique mean and outlier sizes were equally often presented in each single task block. There were no Outlier⁻ displays in the mixed block.

4.2.2 Experiment 1B.

Task details were the same in Experiment 1B except that now Outlier+ displays had the smallest outlier among larger sized circles and Outlier– displays had the equivalent smallest size that was not an outlier (see Figure 4). Accordingly, the largest size estimation task was replaced with the smallest size estimation task. Also, spatial arrangement was reversed for Outlier+ and Outlier– displays in Experiment 1B.

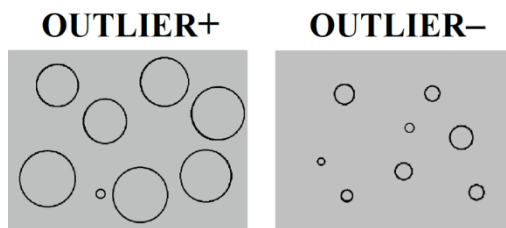


Figure 4. Illustration of displays in Experiment 1B

In the Outlier+ displays, seven member circles randomly sized between diameters of 5.4° and 7.8° . The smallest outlier size was computed by subtracting 5.6° from the local mean of seven member circle diameters. This meant that the outlier circle's diameter was at most 2.25 standard deviations less than the global mean. This also meant that the difference between the diameters of local and global mean was always 0.7° as in Experiment 1B. Possible outlier diameters were 0.8° , 1.0° , 1.2° , and 1.4° whereas possible global mean diameters were 5.7° , 5.9° , 6.1° , and 6.3° .

In the Outlier– displays, the smallest circle sizes were identical to the smallest outlier sizes in Outlier+ displays. Diameters of the other seven circles were randomly chosen from the 13 alternatives as in Outlier+ displays but increasing from the smallest size by 0.2° . Possible mean diameters were 2.0° or 2.2° when the smallest diameters were 0.8° and 1.0° and 2.4° or 2.6° when the smallest diameters were 1.2° and 1.4° .

CHAPTER 5

RESULTS

For each observer, I calculated average absolute error across each condition; I excluded trials in which the absolute error was 3 standard deviations above the average absolute error of the whole sample. Absolute error was the difference between the reported and correct diameters in degrees of visual angle. Because I instructed participants to report the mean size of all circles in a display, the correct mean size value in Outlier⁺ displays was the global mean which includes the outlier. I chose the absolute error instead of error proportional to the correct size as accuracy index² because a key press of a participant changed the test item's radius by one pixel. That was, adjustment of a test item by key presses was fixed but not based on percent change.

Here, I reported results in accordance with the research questions. Note that the design was within-subjects across Experiment 1A and 1B; hence, all analyses were repeated measures.

5.1 The mean size estimation

In order to see whether the presence of an outlier decreases the precision of mean size estimations, I first analyzed the single mean size estimation block across two experiments (see Figure 5). A 2 (experiment: 1A/1B) \times 2 (display: Outlier+/ Outlier-) ANOVA on absolute error turned out a main effect of display, $F(1,25) = 9.19$, $MSE = 1.18$, $\eta_p^2 = .27$, $p = .006$. As expected, mean size reports were less accurate with Outlier+ ($M = .49$, $SD = .012$) than Outlier- ($M = .42$, $SD = .009$) displays. There

² Nevertheless, I ran the analyses also with percent error and the pattern of results were same as the analyses with absolute error in radius.

was no main effect of experiment, $F(1,25) = 3.26$, $MSE = .92$, $\eta_p^2 = .11$, $p = .083$.

However, there was a significant interaction, $F(1,25) = 8.77$, $MSE = 1.52$, $\eta_p^2 = .26$, $p = .007$. This interaction was due to the fact that Outlier– displays of Experiment 1B ($M = .37$, $SD = .011$) had the least amount of error, significantly different than Outlier+ displays of Experiment 1B ($M = .51$, $SD = .014$), Outlier– displays of Experiment 1A ($M = .47$, $SD = .013$), and Outlier+ displays of Experiment 1A ($M = .47$, $SD = .015$). To understand why error was significantly less in Outlier– displays of Experiment 1B, I looked at how characteristics of these displays differed from the remaining ones. Compared to all others, these displays had the smallest mean size ($M = 23$, $SD = 2.24$). As a control, I compared mean reports of mostly small circle sets (Outlier+ of Experiment 1A and Outlier– of Experiment 1B) with mostly large circle sets (Outlier– of Experiment 1A and Outlier+ of Experiment 1B). Mean size estimations were more accurate with mostly small circles ($M = .42$, $SD = .011$) than mostly large circles ($M = .49$, $SD = .011$), $t(25) = 2.96$, $p = .007$, $d = .58$. Thus, the interaction was confounded by the differences in mean size that Outlier+ displays resulted in more error than Outlier– displays in Experiment 1B, $t(25) = 4.1$, $p > .001$, $d = .81$, probably because the former had larger mean sizes ($M = 63.5$, $SD = 2.24$) with which people were already less accurate. Only Outlier+ displays of Experiment 1A consisted of mostly small circles and had smaller mean size ($M = 26$, $SD = 2.24$) as Outlier– displays of Experiment 1B had. When I compared these displays with smaller means, the presence of a larger outlier resulted in greater error, $t(25) = 3.74$, $p = .001$, $d = .75$. Hence, making a small circle an enormous outlier among other smaller circles decreased the precision of mean reports while making a large circle a tiny outlier among other larger circles did not.

One of the main predictions was that the mean size estimations of Outlier+ displays would be biased towards the local rather than global mean. Since I had Outlier+ displays in both the single mean size estimation and the mixed blocks, I calculated bias to local mean in these trials and averaged bias scores in particular conditions for each observer. Specifically, absolute error to the global mean was divided by absolute error to the local mean. Thus, a resulting score above 1 indicated a bias towards local mean and a score below 1 indicated a bias towards global mean. In line with the prediction, bias in mean reports (see Figure 5) was always significantly more than 1 indicating a significant bias towards local mean instead of global mean (all p s < .02).

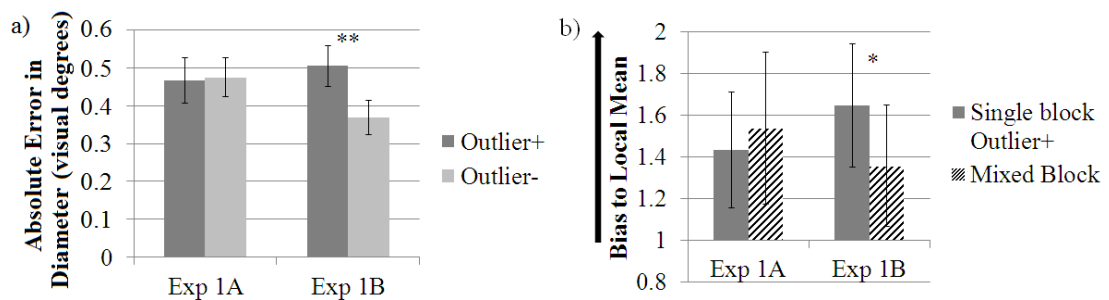


Figure 5. Graphs depicting (a) accuracy and (b) bias towards local over global mean in the mean size reports

Note: * $p = .017$, ** $p < .001$, $N = 26$. Error bars indicate 95% confidence intervals.

5.2 The largest/smallest size estimation

My second question was whether an item is better represented when it is an outlier than when it is not. I conducted a 2 (experiment: 1A/1B) \times 2 (display:

Outlier+/Outlier-) ANOVA on absolute error in the single largest/smallest size estimation blocks (see Figure 6). As predicted, there was a main effect of display,

$F(1,25) = 11.2$, $MSE = .25$, $\eta_p^2 = .31$, $p = .003$. Outlier+ displays ($M = .23$, $SD =$

$.005$) yielded less erroneous largest/smallest size reports than Outlier- displays ($M =$

$.26$, $SD = .008$). Also, the main effect of experiment was significant, $F(1,25) =$

133.69, $MSE = 1.0$, $\eta_p^2 = .84$, $p < .001$. The largest size estimations ($M = .36$, $SD = .011$) were less accurate than the smallest size estimations ($M = .13$, $SD = .004$).

There was no interaction, $F(1,25) = .17$, $MSE = .04$, $\eta_p^2 = .007$, $p = .68$, indicating that both smallest and largest size reports were more accurate when these sizes were outliers of the display.

In addition, I expected that there should be less of a bias towards the global mean when the item size is an outlier than when it is not. Bias was calculated by taking the ratio of absolute error to the difference between the reported and global mean diameters. Therefore, if a bias score exceeds 1, there would be a bias towards global mean instead of correct largest/smallest size. If a bias score is less than 1, there would be a correct bias towards the largest/smallest size over global mean. Then, I averaged bias scores in trials of the particular condition for each observer. In fact, bias towards mean (see Figure 6) was always significantly lower than 1 demonstrating that there was no bias towards mean in all largest and smallest reports (all $ps < .001$). Thus, viewers always correctly biased towards the required size rather than the mean size. However, as expected, outlier reports, regardless of being in the single task or mixed task block, were correctly pulled towards the required size more than non-outlier item estimates (all $ps < .001$, all $ds > 3.5$).

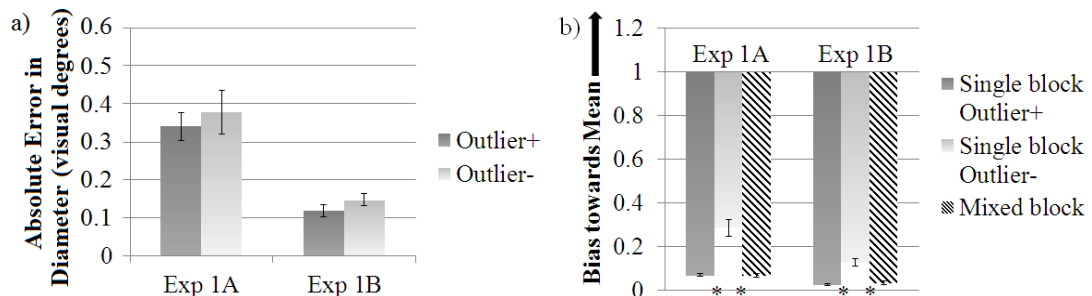


Figure 6. Graphs depicting (a) accuracy and (b) bias towards mean in the largest/smallest size reports

Note: $*p < .001$, $N = 26$. Error bars indicate 95% confidence intervals.

5.3 The outlier and mean size estimation

Last, I asked whether attending to the mean and outlier size concurrently would result in any additional costs. Following this aim, I compared the reports for Outlier+ displays in the single task blocks with the reports in the mixed block which required subjects to focus on the outlier and the mean concurrently. I separately ran a 2 (task: outlier/mean) \times 2 (block: single task/mixed) ANOVA on absolute error for Experiment 1A and Experiment 1B (see Figure 7). In both experiments, there was a main effect of task, 1A: $F(1,25) = 20.02$, $MSE = 1.56$, $\eta_p^2 = .45$, $p < .001$ and 1B: $F(1,25) = 141.87$, $MSE = 2.37$, $\eta_p^2 = .85$, $p < .001$. The outlier estimations (1A: $M = .34$, $SD = .009$ and 1B: $M = .14$, $SD = .006$) were more accurate than the mean estimations (1A: $M = .45$, $SD = .014$ and 1B: $M = .5$, $SD = .015$). Surprisingly, in Experiment 1A, there was a small yet significant benefit of doing both tasks in a mixed fashion, $F(1,25) = 5.04$, $MSE = .37$, $\eta_p^2 = .17$, $p = .034$. For large outlier displays, outlier and mean size estimations were more accurate in the mixed ($M = .38$, $SD = .011$) than single block ($M = .4$, $SD = .01$). Also, non-significant interaction, $F(1,25) = 2.45$, $MSE = .28$, $\eta_p^2 = .09$, $p = .13$, indicated that attending to outlier and mean size concurrently benefited to the precision of both estimations in Experiment 1A.

In Experiment 1B, main effect of block was not significant, $F(1,25) = .63$, $MSE = .62$, $\eta_p^2 = .03$, $p = .44$. There was, nevertheless, an interaction of task with block, $F(1,25) = 5.64$, $MSE = .35$, $\eta_p^2 = .18$, $p = .026$. For small outlier displays, mean size reports were equally accurate in the single ($M = .51$, $SD = .014$) and mixed blocks ($M = .49$, $SD = .018$) which was consistent with the prediction. However, the

smallest outlier reports were more erroneous in the mixed ($M = .16, SD = .008$) than in the single block ($M = .12, SD = .004$), $t(25) = 3.574, p = .001, d = .95$.

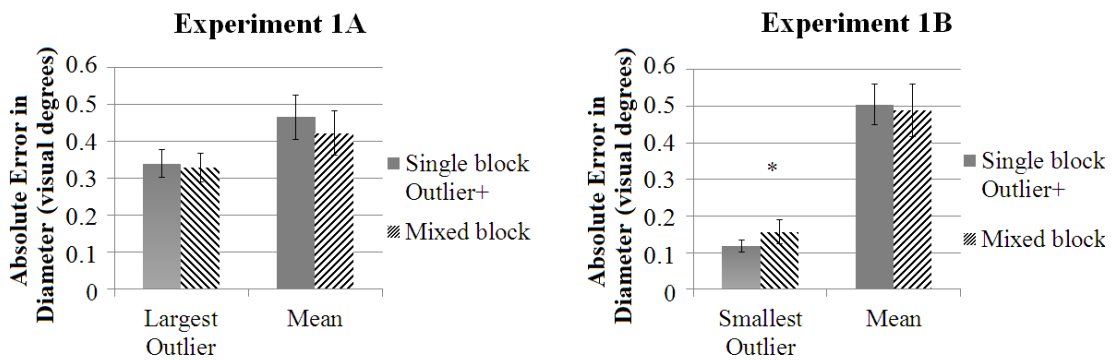


Figure 7. Graphs depicting accuracy for displays with an outlier

Note: $*p < .001$. $N = 26$. Error bars indicate 95 % confidence intervals.

5.4 Estimation biases as a control

Because participants indicated their answers by adjusting a test item, the current paradigm used the method of reproduction. It is known that this method frequently yields asymmetries in estimation performances (Rajamanickam, 2002). Hence, one might predict that people would tend to overestimate smaller sizes while they tend to underestimate larger sizes; this would be a disadvantage of the current method. To address this issue, I calculated an estimation bias score by taking the ratio of the reported to the actual value. Resulting score above 1 indicated an overestimation whereas resulting score below 1 indicated underestimation. There was an overestimation in the smallest item reports only when it was asked in the mixed block ($M = 1.05, SD = .114$), $t(25) = 2.11, p = .045$. On the other hand, there was not any underestimation but an overestimation of the largest item sizes in the single largest size estimation block. The overestimation was evident regardless of whether the largest item was the outlier ($M = 1.01, SD = .026$), $t(25) = 2.66, p = .014$, or not ($M = 1.03, SD = .03$), $t(25) = 4.94, p < .001$, in that single block. Therefore, I showed

that the estimation biases of the method of reproduction were not the case with the present paradigm probably because these biases are sensitive to some specific structural conditions that are absent in the present displays.

CHAPTER 6

DISCUSSION

In this thesis, I investigated the interaction between outlier and ensemble representations. Results suggest the formation of outlier and ensemble representations is a highly intertwined process, which results in high fidelity representations of outliers. Even though there has been substantial work on outlier detection in the visual search literature (e.g. Stuart, Bossomaier, & Johnson, 1993; Treisman & Gormican, 1988), this study is one of the first to directly look at the ensemble and the outlier relationship, using a novel estimation task which specifically provided more direct evidence for representational quality of outliers. I had three major questions in mind; I discuss the relevant findings for each in turn.

The first question I investigated was how the presence of an outlier impacts mean size estimations. I found that mean size estimations of displays that had an outlier were more erroneous, especially when a larger outlier was presented among smaller items. The results are partially in line with the predictions drawn from the model proposed by Rosenholtz and Alvarez (2007). First, the model predicted that more variable displays may lead to greater variance and less precision in mean estimations. Although Outlier⁺ displays did not result in greater variance in estimations than Outlier⁻ displays,³ $F(1,25) = .21$, $MSE = .02$, $\eta_p^2 = .008$, $p = .66$, I nevertheless found that the precision of mean size estimations were significantly decreased when the outlier was a largest item in a smaller set. Even though both conditions similarly increased the variance of the visual display, the findings suggest

³ Variance index was coefficient of variation which was the ratio of *SD* to *M* of a particular condition. The reported statistics was the null main effect of display which was yielded by a 2 (experiment: 1A/1B) × 2 (display: Outlier+/ Outlier-) within-subjects ANOVA. Experiment main effect and the interaction were also not significant, all *ps* > .35.

that there may be factors other than variance that influence the mean size estimation process. The likely candidate may be how large the mean size of a set was to begin with. I found that people estimated larger mean sizes with greater error than smaller mean sizes, replicating earlier work from our lab (Semizer, 2013). Since the sets that had the larger outlier also had small means, any decrease in mean size precision may have been easier to detect. On the other hand, small outliers were always among larger items; the mean size estimations of larger item sets were more erroneous regardless. Thus, the outlier's impact on an already erroneous response may have gone undetected.

Rosenholtz and Alvarez (2007) also predicted that in more variable displays, viewers may average only a subset by discarding items that introduce variability to the distribution. Consistent with this prediction, I demonstrated that the mean size estimations were biased towards the local mean rather than the global mean. This suggests that viewers do not see an outlier as a part of the ensemble and do not incorporate it into the summary statistics. This finding was previously revealed with facial emotions (Haberman & Whitney, 2010). However, I am first to show the dissociation of the outlier from the ensemble with a simple feature, size. Importantly, bias towards local mean was significant for both displays with a larger and a smaller outlier and both in the single task and mixed task blocks. That means viewers tend to omit the outlier from summary statistics regardless of whether the outlier was popping out or not, which is again conceptually consistent with the findings of Haberman and Whitney (2010).

The second question I was interested in was whether outliers were processed more efficiently than non-outliers. Across all trials, I specifically had same sized items that were either an outlier or not. I found two pieces of empirical evidence to

illustrate that an outlier in a visual set is represented efficiently. First, across two experiments, I found that an outlier's size was represented with greater precision than non-outlier items. Second, outlier representations were more veridical than other non-outlier item representations and less influenced from the ensemble representation. I specifically found that outlier representations were closer to the correct size than non-outlier representations. Thus, it may be that outlier representations may be more robust than non-outlier representations. This is in stark contrast to some earlier work suggesting that lower fidelity item representations (albeit non tested outliers) may benefit from the presence of accurate veridical ensemble representations (Boduroglu & Shah, 2014; Brady & Alvarez, 2011; Mutlutürk & Boduroglu, 2014; 2016). This suggests that outlier representations are somewhat unique and not just represented as an item of an ensemble. Specifically, the finding that the outlier representation was not biased towards ensemble representation bolsters this interpretation. In this regard, the findings are similar to those of Huttenlocher et al. (2000). They demonstrated that outliers showed a reduced bias to the center of the frequency distribution and thus were more veridical. They noted that this pattern of findings was driven by one's exposure to the particular item distribution. Hence, efficient and veridical outlier representations could be achieved both by identifying extreme perceptual features, as in the present thesis, and through statistical learning, as in Huttenlocher et al. (2000) study. The veridicality and high precision of the outlier was always the case whatever the outlier size was. Hence, if asked to do so, people can efficiently extract an outlier size and the perceptual saliency of the outlier among other items does not affect this process. This finding is in parallel with visual search literature that there was no difference in

search latencies for small and large circle sizes (with filled circles: Cavanagh, Arguin, & Treisman, 1990; with outlined circles: Stuart et al., 1993).

My third main goal for this particular thesis was about whether ensemble and outlier processing can occur concurrently. There seems to be sufficient evidence to suggest that these may be two very rapid pre-attentive processes that are inter-related and concurrently occurring. In the final mixed block, attending to outlier size along with the ensemble did not result in a particular cost for ensemble representation. Focusing on the larger outlier concurrently with the ensemble increased the precision of mean size estimations. On the other hand, simultaneously attending to the smaller outlier with the ensemble did not affect the precision of mean estimations. However, the bias towards the local mean was reduced for the mean size estimation in the mixed block than in the single mean block in Experiment 1B, $t(25) = 2.56, p = .017, d = .5$. This result is not surprising given that ensemble representations can be formed even when attention is limited by another task (Alvarez & Oliva, 2008) or impaired by neglect syndrome (Yamanashi Leib, Landau, Baek, Chong, & Robertson; 2012). Moreover, this result is in line with our earlier finding that there is no cost for averaging the entire set along with concurrent averaging of the spatial groups constituting the set (Yildirim et al., 2015). Thus, these demonstrate that viewers are capable of averaging the whole set while simultaneously attending to subsets or individual outlier items. This suggests that people can extract the ensemble properties while being capable of perceptually organizing the elements in the display. On the other hand, simultaneous processing of the ensemble along with the outlier yielded different effects on the larger and smaller outlier representations and I discuss this difference below. It is important to note that although I could look at concurrent processing of the outlier and the ensemble, I was not able to speculate on

the time course of these processes. Specifically, the mixed block in the current paradigm did not allowed me to infer any cause and effect relationship between outlier and ensemble extraction. Thus, behavioral modeling can be useful to identify temporal resolution of outlier and ensemble representations. However, behavioral paradigms might be insufficient to explore such fast processes.

One key finding from these experiments is that the fundamental aspects of the outlier may have specific consequences for the representations. Since I manipulated the visual elements on size, I was able to look at differences in results based on the perceptual saliency of the outlier. While the larger outlier in Experiment 1A seem to pop out among small circles, in Experiment 1B, the smaller outlier may have been somewhat obscured by the other large circles in the set. Critically, I found that the presence of a larger or smaller outlier item in ensembles resulted in a different pattern of findings. First, the detrimental effect of an outlier on mean size precision was evident only when a larger outlier was present. Yet, as mentioned earlier this result may have been impacted by the mean size differences across conditions. Second and more importantly, the concurrent processing of the outlier and the ensemble differentially altered the precision of outlier representations. In the mixed block, concurrently processing the mean size and the outlier increased the larger outlier precision whereas it decreased the precision for the smaller outlier. An alternative explanation would be that a smaller outlier might be assimilated by the other large items in the set when concurrently averaging the set. That is, because they were already less accurate with larger sets, people may perceive those displays as a texture. Thus, the texture perception might accelerate averaging in the mixed task which might lead to the absorption of the smaller outlier. These altogether suggest that all outliers may not interact with ensemble processing in the same fashion.

Perceptually non-salient outliers may be less veridically and less independently represented from the ensemble when concurrent processing demands are high.

The present thesis is one of the first steps in clarifying the relationship between outlier and ensemble representations. I found that viewers downplay the outliers in statistical summary representations, outlier items are better represented than non-outliers, and concurrently attending to the outlier along with the ensemble is harmless for statistical summary representations. These results turned out regardless of the perceptual saliency of the outlier. More critically, however, only perceptually salient outliers can be efficiently processed along with ensembles and people concurrently form robust representations of perceptually salient outliers. The findings of this thesis demonstrated that defining an outlier based on distributional characteristics may not guarantee perceptual salience. Future research would need to extend these findings to other visual features and establish whether more complex outliers are processed in a similar fashion. Investigating with more complex visual displays would provide complementary understanding about outliers which can be defined by semantic relationships of an ensemble or a scene.

APPENDIX

INFORMED CONSENT FORM

Bilgilendirilmiş Onam Formu

Arařtırmacıların Adı:

İrem Yıldırım (irem.yldrm8@gmail.com),
Ayşecan Bodurođlu (aysecan.boduroglu@boun.edu.tr)

Adres:

Boğaziçi Üniversitesi, Psikoloji Bölümü, Bebek 34342 İstanbul

Katılacağınız araştırma bir bilişsel psikoloji araştırmasıdır. Bu araştırmanın amacı, bireylerin çok kısa bir süre gördükleri bir görüntüyü nasıl hatırladıklarını incelemektedir. Bu amaçla, bilgisayar başında görsel bellek ile ilgili bir takım görevler yapmanız istenecektir.

Çalışmanın toplam süresi yaklaşık olarak 1 saattir. Deneye katılımınız karşılığında size PSY 101 veya PSY 241 dersinden 1 kredi verilecektir.

Araştırmada tüm kişisel bilgilerin gizliliği esas alınmıştır. Dilediğiniz an araştırmadan geri çekilebilirsiniz.

Sormak istediğiniz bir soru varsa lütfen deneyi yapan araştırma görevlisine sorunuz.

Yukarıdaki bilgileri okudum ve anladım. Anlamadığım kısımlar hakkında araştırmacıdan gerekli bilgileri aldım.

Formun bir örneğini aldım.

Deneye katılmayı kabul ediyorum _____

İsim Soyad:

Öğrenci no:

Tarih:

Deney Yürütücüsü:

Katılımcı İmza:

Deney Yürütücüsü İmza:

REFERENCES

- Albrecht, A. R. & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.
- 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.
- Attarha, M. & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, *77*, 1116–1131.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.
- Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, *59*, 171–186.
- Boduroglu, A. & Shah, P. (2014). Configural representations in spatial working memory. *Visual Cognition*, *22*(1), 102–124.
- Brady, T. F. & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392.
- Brady, T. F. & Tenenbaum, J. B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 411–416). Austin, TX: Cognitive Science.
- Brady, T. F. & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*, 85–109.
- Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 479–491.
- Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception and Psychophysics*, *70*, 1327–1334.
- Chong, S. C. & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.

- Chong, S. C. & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, 45(7), 891–900.
- Corbett, J. E. & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, 138, 289–301.
- de Fockert, J. W. & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, 62(9), 1716–1722.
- Haberman, J. & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17, 751–753.
- Haberman, J. & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734.
- Haberman, J. & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception Psychophysics*, 72, 1825–1838.
- Haberman, J. & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York: Oxford University Press.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, 129, 220–241.
- Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203.
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, 142(2), 245–250.
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11):9, 1–8. doi:10.1167/8.11.9
- Mutlutürk, A. & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, & Psychophysics*, 76(8), 2276–2285.

- Mutlutürk, A. & Boduroglu, A. (2016). *Configural representations summarized into the centroid*. Manuscript in preparation.
- 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. & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.
- Oliva, A. & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*, 520–527.
- 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*(1), 29–46.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- Rajamanickam, M. (2002). *Modern psychophysical and scaling methods and experimentation*. New Delhi, India: Concept Publishing Company.
- Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*, 3157–3163.
- Rosenholtz, R. & Alvarez, G. A. (2007). How and why we perceive sets: What does modeling tell us? [Abstract]. *Perception*, *36*, 8. Retrieved from <http://www.perceptionweb.com/ecvp/ecvp07.pdf>
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17. doi:10.1167/12.4.14
- Semizer, Y. (2013). *Representation of similar objects as a group: Effects of variance on ensemble representations*. Bogazici University, Istanbul, Turkey.
- Stuart, G. W., Bossomaier, T. R., & Johnson, S. (1993). Preattentive processing of object size: implications for theories of size perception. *Perception*, *22*(10), 1175–1193.
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 329–337.

- Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, 78(3), 392–402.
- Torralba, A. & Oliva, A. (2003) Statistics of natural image categories. *Network: Comput. Neural Systems*, 14, 391–412.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194–214.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40A, 201–237.
- Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95(1), 15–48.
- Whiting, B. F. & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, 18, 484–489.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Yamanashi Leib, A., Landau, A. N., Baek, Y., Chong, S. C., & Robertson, L. (2012). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. *Frontiers in Human Neuroscience*, 6:267, 1–11, doi:10.3389/fnhum.2012.00267
- Yildirim, I., Ögreden, O., & Boduroglu, A. (2015). *Impact of spatial grouping on mean size estimation*. Manuscript submitted for publication.