THE EFFECT OF INTERACTION THROUGH GAZE-CUE SHARING ON DECISION MAKING PROCESS: AN EXPERIMENTAL INVESTIGATION OF KEYNESIAN BEAUTY CONTEST

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMRE ERÇİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COGNITIVE SCIENCE

SEPTEMBER 2019

THE EFFECT OF INTERACTION THROUGH GAZE-CUE SHARING ON DECISION MAKING PROCESS: AN EXPERIMENTAL INVESTIGATION OF KEYNESIAN BEAUTY CONTEST

submitted by **EMRE ERÇİN** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin _____________

Dean, **Graduate School of Informatics**

Prof. Dr. Hüseyin Cem Bozşahin _____________

Head of Department, **Cognitive Science**

Assist. Prof. Dr. Murat Perit Çakır

Supervisor, **Cognitive Science, METU** _____________

Examining Committee Members:

Assoc. Prof. Dr. Cengiz Acartürk _____________

Cognitive Science Department, METU

Assist. Prof. Dr. Murat Perit Çakır _____________

Cognitive Science Department, METU

Assist. Prof. Dr. Murat Ulubay _____________

Department of Management, YBU

Date: 13 September 2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: EMRE ERÇİN

Signature :

ABSTRACT

THE EFFECT OF INTERACTION THROUGH GAZE-CUE SHARING ON DECISION MAKING PROCESS: AN EXPERIMENTAL INVESTIGATION OF KEYNESIAN BEAUTY CONTEST

Erçin, Emre

M.S., Department of Cognitive Science

Supervisor : Assist. Prof. Dr. Murat Perit Çakır

September 2019, 75 pages

The present study investigates the effect of group gaze-cue interaction on the decisions, decision times and eye-movements of the participants in the game of beauty contest. The experiment is designed in the Group-Eye Tracking (GET) platform to create different gaze-cue sharing conditions, ranging from no gaze-cue to all gaze-cue conditions where each participant's gaze location is visualized on all other participants' screens. It is hypothesized that the anchoring effect, which is argued to guide the decisions in the beauty contest task, would be influenced by the increasing awareness of the participants regarding each others' eye movements preceding their decisions. The results indicated that there was no gaze-cue sharing effect on the eye movement patterns and no significant difference was observed among the gaze-cue conditions for the decided value derived from individual choices. However, for all contest types it was found that decision times are significantly higher in the all gaze-cue condition. The indifference observed for guesses and eye-movements can be partly accounted by the reduced alertness due to repeated rounds of play. Higher decision times in the all gaze-cue condition might be due to the deception strategies that emerged and the dynamic re-anchoring of the "satisficing" guess according to other players' behaviors revealed through their gaze-cues.

Keywords: Beauty Contest, Anchoring, Group Eye Tracking, Decision Making, Theory of Mind

ÖZ

GÖZ-İZİ YOLUYLA ETKİLEŞİMİN KARAR VERME SÜREÇLERİ ÜZERINE ETKISI: KEYNES GÜZELLİK YARIŞMASININ DENEYSEL BIR İNCELEMESİ

Erçin, Emre

Yüksek Lisans, Bilisşel Bilimler Bölümü

Tez Yöneticisi : Dr. Öğretim Üyesi Murat Perit Çakır

Eylül 2019 , 75 sayfa

Bu çalışma grup göz-izi etkileşiminin kişilerin güzellik yarışması oyunundaki kararları, karar verme süreleri ve göz hareketleri üzerindeki etkisini araştırmaktadır. Deney, hiç göz-izi olmaması durumundan her katılımcının bakış konumunun diğer katılımcıların ekranında canlandırıldığı tam göz-izi koşulları arasında değişen farklı göz-izi paylaşım koşulları yaratmak amacıyla Grup Göz Izleme (GET) platformunda tasarlanmıştır. Güzellik yarışması oyunundaki kararları yönlendirdiği düşünülen çapa atma etkisinin, katılımcıların birbirlerinin göz hareketlerine yönelik farkındalıklarının artmasından etkileneceği beklenmektedir. Göz-izi paylaşımının göz hareketleri örüntüsü üzerinde bir etkisi olmadığı ve bireylerin tercihlerine dayanarak verdikleri değer kararlarının göz-izi koşulları arasında farklılaşmadığı bulunmuştur. Ama, katılımcılara sunulan bütün yarışma türleri için karar süresinin anlamlı bir şekilde farklılaştığı görülmüştür. Tahminler ve göz hareketlerinde fark bulunmaması, tekrarlı oyundan doğan tetiktelik seviyesinin azalmasıyla kısmen açıklanabilir. Bütün göz-izi kosullarında daha uzun karar verme süresinin olması kandırma stratejilerinden ve diğer oyuncuların göz-izi yoluyla ortaya çıkan davranışlarına göre "tatmin edici" tahminin dinamik yeniden çapa atmadan kaynaklanmış olabilir.

Anahtar Kelimeler: Güzellik Yarışması, Çapa Atma, Grup Göz İzleme, Zihin Kuramı, Karar Verme

to my family and friends

ACKNOWLEDGMENTS

First of all, I would like to express my highest gratitude to my supervisor, Assist. Prof. Dr. Murat Perit Çakır for his benevolent patience, understanding, encouragement, optimism and support for three years.

Besides my supervisor, I owe a debt of gratitudes to Assoc. Prof. Dr. Cengiz Acartürk for all his trust, care, motivation and resource and opportunity provided.

I would like to thank the entire METU Cognitive Science department for being a humble and generous nest provided to knowledge-seekers, intellectual candidates. I would like to especially thank to Assist. Prof. Dr. Umut Özge for his warm welcome to department and naively caring and feeding the curiosity of mine.

For partially supporting the present thesis via the project 116E570 "Multiuser Eye Tracking Platform for Social Gaze" (2017, 2019), I would like to thank to Scientific and Technological Research Council of Turkey (TUBITAK)

I would like to thank Mine Özkul, Mani Tajaddini, Efecan Yılmaz and Çağatay Taşçı for their assistance and help for the computational design of the experiment. Also, I would like to thank to Hazal Arpacı and Şura Genç for their assistance throughout the conduction of the experiment.

I thank to my friends Hatice Gonca Bulur, Gizem Özen and Mehmet Akif Berber for their friendship, care, support and encouragement. I also would like to thank to Ela Altın for being there for me all the time. Lastly, I want to thank to Şeyma Uzunöz for her companionship for years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

TUBITAK	Scientific and Technological Research Council of Turkey
NE	Nash Equilibrium
GET	Group Eye Tracking
RS	Research Question
H	Hypothesis
ms	millisecond
cm	centimeters
ANOVA	Analysis of Variance
px	pixel

CHAPTER 1

INTRODUCTION

1.1 Purpose of the Study

The world hosts a trembling balance between the order and chaos, asynchronous and simultaneous, hypothetical and actual, cool and hot. Thousands of birds flying, fish swimming, ants walking, horses running and people living together follows the same basic rules emerging from a pattern called life. Even though the outcome seems similar, from the right angles, the complexity of the mechanism can be differed between organisms. Humans, as many other living creatures in the earth, tend to construct social interactions to survive. Even the most basic interactions require prediction of behavior by a simultaneous reasoning of observing or guessing other’s behaviours. Heuristics of reasoning about other’s cognitive processes and mental states necessitate a theory of mind existing in the organism. Underlying mechanism of the strategic reasoning can be reverse engineered to investigate the underlying mechanism of human mind as well.

Game theory (Neumann & Morgenstern, 1944) is a mathematical framework to investigate the strategic interactions between competing or cooperating players. As the game theoretical concept of Nash Equilibrium (NE) (Nash, 1950) suggests, there is always an optimal outcome, a state of balance that holds no necessity for the search of a new strategy. It is expected that a rational decision maker is going to reach the equilibrium over time. Yet there are games that the concept of NE and the rational decision-maker cannot be easily integrated. Guess the average (Moulin, 1986) or in other words Beauty Contest (Nagel, 1995) is a game where the satisficing reasoning can be more rewarding than optimizing reasoning. Cutting off the depth of reasoning intuitively is more rewarding than reaching the NE (Bosch-Domènech, Montalvo, Nagel & Satorra, 2002). This deviation indicates rational decision makers’ path of reasoning follows "satisficing" rather than "optimizing" due to the limits of the mental capacity (Simon, 1996). To reach a point which can be called as a decision of an individual, there are mental shortcuts for reasoning called heuristics. Anchoring to an externally provided or self-generated value while strategical reasoning is one of the heuristics which can also be called as a cognitive bias of in low elaboration thinking process (Tversky and Kahneman, 1974; Wegener et al., 2000).

This research investigates the difference in the strategical reasoning caused by the exposure of the state of other participants while playing a "guess the average" game. There are three different exposure methods; no gaze-cue sharing, self gaze-cue sharing and all gaze-cue sharing. The deviation expected to be seen is the change of the guesses, guess times and eye movements.

1.2 Research Question and Hypotheses

In this section a research question and hypotheses are presented. Research question interrogates gaze-cue sharing conditions effect on reasoning process. Five hypotheses emerged from research questions expects deviation on the guess results, guess times and eye-movements.

1.2.1 Research Question

Depth of reasoning, level-k model, is developed for the beauty contest game to explain the reasoning process and significantly higher frequency for the specific numbers in the interval given to the player. This several numbers are changing according the self-generated and externally provided reference points. Anchoring heuristics examines the change of the final judgment according to the initial reasoning point which refers to reference point in the depth of reasoning model. There are three main experimental conditions defined for the task of beauty contest game. *no gaze-cue sharing*, *self gaze-cue sharing* and *all gaze-cue sharing* conditions are exposed to the participants repeatedly in an experiment. *all gaze-cue sharing* deviates the anchoring bias by decreasing the uncertainty, increasing the cognitive load and providing simultaneous giving information about other players' guessing procedure. Hence, it is expected that there are going to be different strategical reasoning for this particular condition. To reduce the difference of whether gazing an empty ruler (no gaze-cue sharing) getting information about self-including all participants (all gaze-cue sharing) there is also a condition where a participant only takes an information about their own gaze-cue (self gaze-cue sharing). All the information gathering is made with a ball-shaped indicator moving on the ruler according to gaze of the participants. In order to measure the deviation in the decision process, how much time spent on the process and the outcome of the process is going to be measured. For the physiological effects, eye movements are going to be detected measured and compared. The study aims to investigate three research questions:

1. Do gaze-cue sharing conditions have an effect on the reasoning process of the players?

1.2.2 Hypotheses

Research questions stated above can be defined as hypotheses. A scientific proposition requires an operational definition for the proper questions. Research Question (RS) can be converted into the Hypothesis 1 (H_1) as the difference expected the guess made within gaze-cue conditions. RS in the same pattern can be formulated as H_2 as the difference expected the guess time within gaze-cue conditions. To measure the alertness, eye tracker devices is going to be used. There are three measures of alertness is chosen: average fixation duration, average saccade amplitude and saccade frequency. Therefore, RS is going to be allocated into three more hypotheses H_3 represents the difference in average fixation duration, H_4 refers to difference in average saccade frequency and H_5 refers to difference average saccade amplitude. Hence five different hypotheses can be defined:

1. There is an effect of gaze-cue sharing condition on guesses of players
2. There is an effect of gaze-cue sharing condition on guess times of players
3. There is an effect of gaze-cue sharing condition on average fixation duration of players
4. There is an effect of gaze-cue sharing condition on saccade frequency of players
5. There is an effect of gaze-cue sharing condition on saccadic amplitude of players

1.3 Organization

The thesis is including five chapters. The first chapter is the introduction chapter presents; "Purpose of the Study", "Research Question and Hypothesis" and "Organization". Second chapter tells about the background and the literature review. In second chapter cognitive process of the anchoring heuristics, theory of mind and social cognition, beauty contest game and it’s variations in the literature and lastly eye tracking studies presented. The third chapter is including the methodology and the materials used in the study. Participants, design, procedure and data preparation is presented. Also group eye tracking platform which is used in the study is taken at hand. The fourth chapter demonstrates the results of the study. Results are including the difference of guesses, guesses times, saccade amplitude, saccade frequency and fixation duration. The fifth chapter includes the discussion of the results, conclusion, limits of the study and future directions.

CHAPTER 2

BACKGROUND

2.1 Beauty Contest Game

This section covers a brief history of the Beauty Contest game, the original game and its variations. Present study is inherited from the original study but influenced by the variations to adapt the environment.

2.1.1 A Brief History of the Beauty Contest

Game Theory is a branch of mathematics aiming to model an agent's best action in specific conditions which is developed by Neumann and Morgenstern (Neumann & Morgenstern, 1944). The concept of Beauty Contest as a game is firstly developed by John Maynard Keynes (Keynes, 1937) in his book of *The General Theory of Employment, Interest and Money.* Keynes gives an example of a "Beauty Contest" while explaining the stock market professional investment. The analogy of "Beauty Contest" comes from newspaper competitions to find out the six prettiest faces around hundreds of photographs. The task is not to find the most beautiful face for a competitor but to guess the average preferences of all competitors.

> Or, to change the metaphor slightly, professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees. (Keynes, 1937, p. 96)

Herve Moulin is the first person who integrated the concept of "Beauty Contest" to the literature of game theory with the name of "Guess the Average" (Moulin, 1986). Moulin gave the example of guess the average game under the "Successive Elimination and Sophisticated Equilibrium" title as a dominance solvable game. The game expected from its competitors to guess a number between 1 and 999 while giving attention to a given "P" value. The game is shown as the first and the clearest example of its kind hence any number under the multiplication of 999 and p is dominant over other guesses. The iteration of the multiplication of max dominant guess and p value brings out a singleton set of dominant value hence a sophisticated equilibrium.

Although not being the inventor as she claims (Bühren, Frank, & Rosemarie, 2012), the modern version of the game is developed by Rosemarie Nagel (Nagel, 1993, 1995). Nagel's studies on the subject is the first recorded experimental study on the subject. Nagel's version of the game is very similar, nearly identical with Moulin's version. The only difference between two is the restrictions and the p values given to players. In Nagel's version limit of the guess for the competitors is between 0 and 100. In Moulin's version of the game p value is exampled as "2/3" but Nagel conducted the experiment with two additional p values "1/2 and 4/3" (Nagel, 1995).

Nagel's first encounter of the game was based on Herve Moulin's example (Bühren et al., 2012). However, she intended to name the game as Beauty Contest from the inspiration of Keynes. On the other hand, Keynesian Beauty Contest game has no p value indicated. Hence, later the game also called as P-Beauty Contest (Ho, Camerer, & Weigelt, 1998).

2.1.2 The Original Beauty Contest and Variations

Keynes brought the concept of the beauty concept and Moulin modified the concept. After that Nagel tells that she played the game as an economics class thought experiment (Bühren et al., 2012). First recorded experimental study on the game was recorded in 1993 by Nagel herself (Nagel, 1993, 1995). Her study was conducted with 15-18 participants per session. Sessions were held in classrooms. A session consists of four rounds of game without any time limit. Researchers expected from participants to guess a number between 0 and 100. A participant who guessed the number closest to average of the guesses - multiplied with "p value" was the winner of the round. p value was written on the board. A participant is only exposed to only one p value. After each round, participants were given feedback for the all chosen numbers, mean of the chosen numbers and the winner number. The aim of the study was to understand the depth of reasoning, modeling the rational behavior, learning with feedback between rounds. It is found that although first rounds do not show any, later rounds iteratively show steep-curve of converging to Nash Equilibrium.

There have been several different versions of the Beauty Contest game after Nagel's research (Duffy & Nagel, 1997; Ho et al., 1998; Bosch-Domènech et al., 2002; Weber, 2003; Slonim, 2005; Kocher & Sutter, 2005, 2006). All of the variations are inspect different features of the games.

Duffy and Nagel set different calculation of the target number for the players. Instead of using only mean, in order to test the robustness, researches also used median and the maximum number target calculation systems. The only "p Value" used was 1/2 for all sessions. It is found out that participants' pay less attention to extreme guesses respectively from mean treatments to median treatments and lastly maximum guess treatments. When compared by the convergence to equilibrium percentage between rounds; the order from greater to least is median, maximum and mean. However, no substantial difference is found average initial choice which was around "36" (two levels of reasoning) in the original game (Duffy & Nagel, 1997).

Ho, Weigelt and Camerer (1998) developed an iterated best response model of the game with varied version of the game. The experiment conducted with two different conditions for reasoning thresholds: Infinite Threshold and Finite Threshold. The difference between the infinite and finite threshold is being provided by changing the limits. For the infinite threshold, it is expected from competitors to approach zero for with infinite reasoning steps, so the limit is set between 0 to 100. However, if the limit is set to 100 and 200 and if the p value is set more than "1", than it is expected from competitors to reach the unique equilibrium in several reasoning steps. Also group size conditions are defined with three-competitor and seven-competitor conditions. Lastly, the difference between inexperienced and experienced players are compared according to their performance. The operational definition for experienced player is defined as a player who played the game of beauty contest or a *similar* game for at least one time. Different p values (*0.7, 0.9, 1.1, 1.3 and 3*) is used for equally distributed ways. The game is played 20 rounds while the first and second 10 rounds are different for the guess limit (threshold condition type) and p value. It is found that all the variables affect choices and learning in different ways. Finite threshold condition and p values more than "1" increases the convergence percentage through equilibrium and larger groups provides closer choices to equilibrium as well. It is seen that for the first round there is no difference between experienced and inexperienced competitors and occasionally confused guesses (spoilers) can be seen (Ho et al., 1998).

Keynes presented Beauty Contest as a newspaper experiment played as a one-shot game between thousands of participants (Keynes, 1937). In 2002, Bosch-Domenech et al. (2002) conducted a one-shot newspaper experiment loyal to Keynesian Beauty Contest. Experiment was conducted in three different countries: USA (conducted by Thaler in *Financial Times*), Spain (conducted by Bosch and Nagel in *Expansion*) and Germany (conducted by Selten and Nagel in *Spektrum der Wissenschaft*). One-shot games are played with p value of "2/3" and only Spain has guess restriction between 1 and 100. The others are between 0 and 100. Time limit for answers were one week for Spain and two weeks for America and Germany. The importance of the study comes from self-explanation made by competitors. While there was no difference in distributions of the total guesses made in every country (spikes at 33, 22 and 0), researchers clustered six different strategies from the feedback made by participants:

- **1. Fixed Point**: This strategy occurs when a participant give a Nash equilibrium based logical answer.
- **2. Iterated Elimination of Weakly Dominated Strategies Plus Rounding, Trembling and Other Rules of Thumb**: This strategy uses the iterated dominance strategy (Ho et al., 1998) however knows that not all participants are going to act rationally and according to Nash equilibrium. Hence, they cut the strategy earlier before reaching the logical answer. After finding a set of guesses to make, a competitor chooses one according to their belief about the other competitors.
- **3*a*. Iterated Best Reply - degenerate *Level* - ∞ Plus Rounding, Trembling and Other Rules of Thumb**: Degeneration defines that competitors who uses this strategy believes that they are thinking one or more level deeper than the rest of the competitors. Difference between Iterated Best Reply and Iterated Dominated Strategy is that Iterated Best Reply reaches an exact number rather than a set of dominant numbers (Ho et al., 1998). *Level* - ∞ refers to a participant asymptotically reaches to "zero" and modified their answer according to their belief about other competitors.
- **3*b*. Iterated Best Reply - degenerate *Level - 1* Plus Rounding, Trembling and Other Rules of Thumb**: This strategy is nearly as same as 3*a* however it only uses the iterated best response strategy for once and modify.
- **3*c*. Iterated Best Reply - degenerate *Level - 0***: *Level - 0* refers to not using the iterated best reply strategy and choosing a random number in the set of numbers presented by researchers. In this case a participant does not have to have a motivation to win an argument and choose a number of "42" because of a novel that they are fan of such as *Hitchiker's Guide to the Galaxy* (Adams, 1995).

- **4. Iterated Best Reply - non - degenerate**: Non-degenerate strategies consider a participant can be ahead or the behind of the participants themselves in terms of level of reasoning.
- **5. Experimenter**: Experimenter strategy is unique to newspaper experiments. A participant applies this strategy gather as many people as they can and conduct the same experiment with them. In the end they go one level of reasoning further from the outcome of the experiment to win the game.
- **6. Group Decision-Making**: This is nearly identical to *experimenter* strategy and it cannot be used in the laboratory experiments. The difference is that participant does not uses a non-degenerate approach to the outcome of the experiment. Participant, on the other hand, uses exactly same outcome of the experiment conducted by themselves to represent a non-degenerate approach. It is found that *experimenter* and *group decision-making* strategy predict the target number better than the rest of the strategies (theorist approaches) (Bosch-Domènech et al., 2002).

It is found that %80 percent of the participants used the iterated best reply non-degenerate strategy and %15 of the participants used iterated best reply *level-0* strategy (random choice). However, although %81 of participants describes their strategy with stating Nash equilibrium, they guessed higher than the equilibrium (Bosch-Domènech et al., 2002).

Weber (2003) focused feedback and priming effect in a different version of the original study. The game is played with a p value of "2/3". Besides a control group, three experimental group is used: In no-feedback-no priming condition, as stated no feedback is used and participants were only informed by question number. In no-feedback-low-priming condition, low priming is made by informing participants by the process of the calculating the scores but not giving any information about the scores and the winners. no-feedback-high-priming condition, on the other, expected from participants to guess the average (not the average multiplied with p value) after each round to force the subjects to think about the game. "Average guessing" in high priming condition did not affect the official scores of the game. It is found that, convergence through equilibrium can be observed without feedback however there is no significant support for the difference between varying priming conditions (Weber, 2003).

Slonim version of the game aims to inspect the experience factor of the participants (Slonim, 2005). "Median-of-the-guesses" and "2/3 p-value" is used for this version and experimental conditions consist of the "same experience" and "mixed experience" experimental groups. The division between experience level is being provided by the procedure of the game. All games are played with three participants with three games and three rounds each game. A participant in same experience group played three games and nine rounds in total with a permutation of nine players. On the other hand, there are one insider and three outsider groups for mixed experience condition. Insider group played three games and for each game insider group played the game with an outsider group which has not played the game before. Hence, the insider group of players played three games, with a same non-experienced group, three rounds less experienced than the insider group and six rounds less experienced than the insider group. It must be noted that outsider group participants is knowing that they are playing with an experienced insider group. It is found that, all non-experienced played in the same way whether they played with an experienced or non-experienced group. However, players learn to adjust their guesses in the next rounds based on the experience of the players they were playing with. On the other hand, experienced players reconditioned the convergence through Nash equilibrium when they are encountered with non-experienced players to win the game. In the end, experienced players earned more reward than the non-experienced players (Slonim, 2005).

Kocher and Stutter (2005) studied the type of decision makers to investigate the economical decision making. The group and individual differences have been looked up with comparison of the original version of the game. 17 individual participants and the 17 groups of three participants compete in 4 session games. It is found that groups are not making "smarter" decisions but convergence to Nash equilibrium is earlier than the individuals (Kocher & Sutter, 2005).

The effect of time pressure on the decision making process during the Beauty Contest game has also been investigated (Kocher & Sutter, 2006). Three groups are designed to represent longer sessions (120 seconds), shorter sessions (15 seconds) and incentive sessions (faster decisions give higher payoffs with 15 seconds limit). The game is played for 24 rounds with p values "2/3", "2/5" and "1/5". There is also a constant is given to players to add to the native average before multiplying with the p value to increase the complexity of the game. This process creates an interior equilibrium. To make an interion equilibrium, a for each trial a constant number is given to the players. Hence instead of reaching "0" they had to make their calculation they have to take this number into account. Between longer and shorter sessions, it is found that equilibrium convergence and the amount of payoffs decreases. On the other hand, quickness incentive condition does not reduce the quality of the decision making when compared to short sessions (Kocher & Sutter, 2006).

The usage of mouse tracking (Costa-Gomes, Crawford, & Broseta, 2001) and eye tracking devices on game theory and decision sciences studies enhanced the understanding of hypothetical reasoning process (Müller & Schwieren, 2011; Chen, Huang, & yi Wang, 2018). Costa-Gomes and Crawford (2001) conducted a mouse tracking study to track back the decision process. The study constructed as a two-person beauty contest game using MouseLab. It is found that the level - k reasoning processes are observable underlying mouse tracking (Costa-Gomes et al., 2001).

An eye-tracking study for beauty contest is made using a ruler in the screen and instructing players to look at the number that they are guessing. The Study is done gathering participants one by one into the room and then merging the data of ten consecutive participants. There was no information was given between rounds. The behavioral data could have been replicated but it is found that people are strategically sophisticated reaching their decisions. So, no linear algorithm can be found under the decision process (Müller & Schwieren, 2011). There is also a spatial beauty contest game developed to use a two dimensional plane for the visual search and reasoning process tracked by eye movements while the p values are the replaced with asymmetric address indicators about where to look in the screen, the two person game is more successful modeling the decision process of the individuals using eye trackers (Chen et al., 2018).

The games briefly presented above holds different features tweaked for the original study. To make a more sterile investigation over the RS, some of the features are been taken from the studies presented. The general features of study such as p values and trial repetition numbers are taken from the Nagel’s original study (Nagel, 1995). Studies made clear inference over the rules of competitions and conventional "average" method is chosen to see the effect of gaze-cue sharing (Ho et al., 1998; Slonim, 2005). Score page is directly taken from what has been given in the original study (Nagel, 1995) on the other hand the standardization of the priming is being taken under control after the results of the Weber’s study (Weber, 2003). Kocher and Stutter’s studies gave an insight on not using time pressure on the players and providing no coordination between group of players (Kocher & Sutter, 2005, 2006). A basic ruler on the screen method is adapted from Mueller and Schieweren’s study to represent the original study in a visual way (Müller & Schwieren, 2011). The reason of the spatial beauty contest is not used that the group size of the planned study as five. Also, the asynchronic nature of the spatial beauty contest was not serving the essence of the *RS* (Chen et al., 2018).

The current study used eye-tracking method with the aims of creating the mechanism of the gaze-cue sharing condition and identifying the eye-movements of the participants. Hence, the scope of the study is being shifted to cognitive science from micro-economics. The strategic reasoning of the participant expected from the participants is tried to be deviated using continuous information from other participants. While the game is still can be won with NE, participants' behavioral deviation because of the "hot" competition in their gaze-cue sharing condition is the novel research approach to the subject.

2.2 Depth of Reasoning

NE is a state where none of players has to change their actions in order to utilize their reward. Beauty contest is one of the games that the rule of Nash Equilibrium (*NE*) of the game for *n* players and [0, 100] guessing restrictions is:

If g_i is the guess made by player i , for i=1,2,…,n , the payoff, $P_j\ (g_1, g_2, \ldots g_n)$, to player j is given by

$$NE(p) = \begin{cases} 0, & p < 0 \\ 100, & p > 1 \end{cases}$$

However, playing equilibrium does not bring out the optimized reply (Nagel, 1995; Bosch-Domènech et al., 2002). As with the feedback between trials, in repetitive games people converge to the NE of the game. However winning number never becomes the equilibrium. In order to explain the deviation from this utilizing economic rationality, level-k reasoning model is developed by Nagel. Level-k reasoning is actually a model eliminates the weakly-dominant answers as Moulin suggest. Moulin took "100" as the initial reference point to refer to the first step of reasoning by elimination of the weak answers (Moulin, 1986). Nagel advances the initial reference point to "50". This advancement lets the "p" values more than one can be calculated with level - k reasoning. Initial reference point only refers to first trial. For the rest of the trials, average guess of the last trial becomes the reference point for calculation (Nagel, 1995). Hence the the level - k reasoning is, while reference point is *R*, level of reasoning is *k* and the interval for the expected guess for step-k reasoning (*t*), given below:

$$R^* p^{k-1/4} < (t) < R^* p^{k+1/4}$$

It is found that the answers given to the trials are clustered into the boundaries of the *k* steps. While the infinite iteration approaches equilibrium asymptotically, the answers most frequent into the "step 2" reasoning as *Figure 1.1* demonstrates. The heuristics beyond the level - k reasoning assumes that they are always one step further than the rest of the players (Ho et al., 1998; Camerer et al., 2004). If a player is reasoning about guessing "step 1" then she assumes that the rest of the players are going to guess in average step 0. In this case, a player using level - k reasoning is a "predictive" player and assumes that either other players are less predictive or myopic. Step 2, for the majority of the participants seems satisficing. Players tend to make satisficing decisions because of the bounded rationality theory rather than optimizing decision which refers to economic rationality (Simon, 1996).

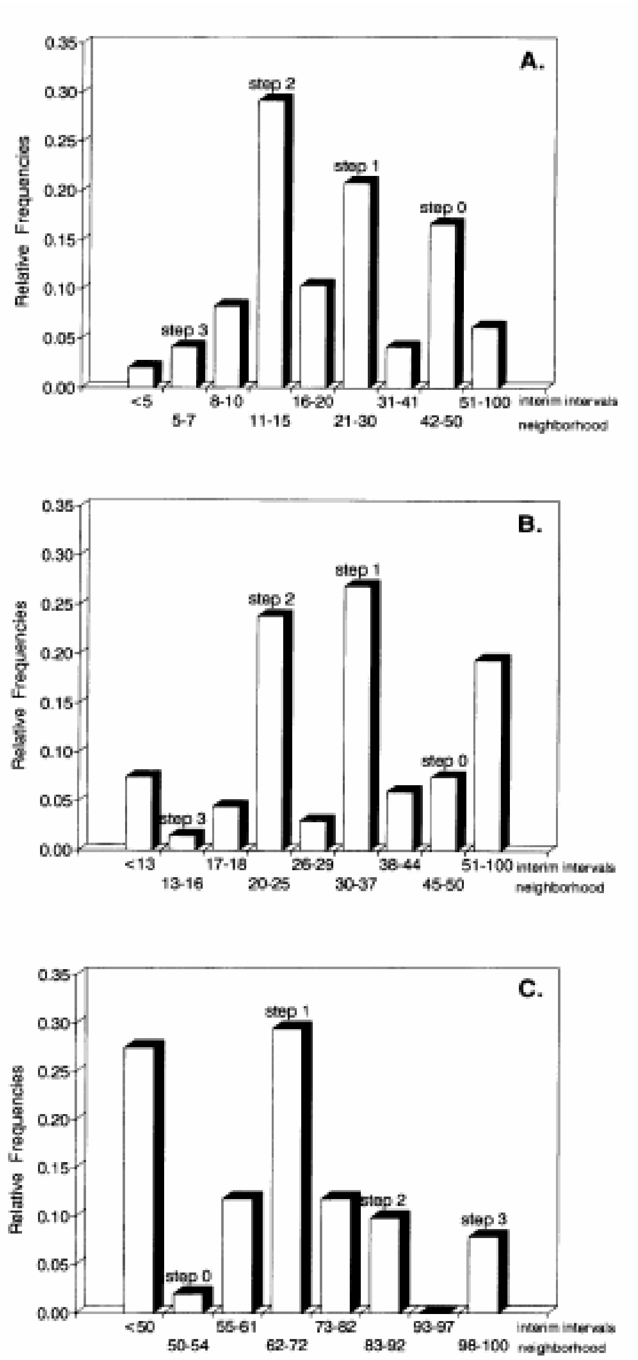


Figure 2.1: Frequency of guesses separated by depth of reasoning model (Nagel, 1995, p. 1317). Frequency bars belongs to "1/2", "2/3" and "4/3" p values respectively

2.3 Social Cognition and Theory of Mind

Social cognition branch of research focuses on the cognitive process of interactions between members of a group (Forgas, 1983). The mechanism behind social cognition is based on the information gathering and sharing to increase the survival rate in group (Frith & Frith, 2012). Social decision making in this sense needs understanding the other's emotions and reason. While understanding emotions requires empathy, understanding reasons requires theory of mind (Frith & Singer, 2008). For these two different concepts of understanding, two different ways of ways of thinking processes are claimed to exist in the human mind.

Theory of mind is one's ability to attribute mental states such as beliefs, desires and intentions to other ones. Theory of mind as an ability is gradually demonstrated at the age of four (Perner, Leekam, & Wimmer, 1987) and is acquired between at the age of 6 or 7 (Perner & Wimmer, 1985). This cognitive evolution has been an interesting subject for psychologists and biologists for decades. Most of the studies are focused on the emerging process (Apperly, 2010).

There are also studies about how a prediction about others changes the strategy of action. Backward induction supposedly used in the game theory. Two person games on prediction on the other hand shows that players are using backward induction and not using the optimal strategies. Meijering et al. (2013) followed the directions of fixation in a "marble drop" game and defined out the forward reasoning rather than backward reasoning is used in the game. This result shows that people, while doing a theory of mind tasks, are using more likely casual reasoning rather than economically rational reasoning. Hedden and Zhang (2002) developed the terms of "myopic" and "predictive" to categorize the player types based on their reasoning in the game. Predictive gamers are the ones who are adjusting their actions based on the other players anticipated actions. "Myopic" players are utilizing the optimal strategy without concerning about others' action. It is found that dependent on reasoning about other's actions, players tend to act predictive and assuming other players are myopic. Goodie et al. (2010) studied when players are "predictive" and when they are "myopic". It is found that in competition in the game contributes to using higher order reasoning about others' actions, while simplicity is not a significant factor in the game. Therefore, beauty contest is game where it is expected that participants use higher order reasoning. Studies shows that participants are either using level-1 or using level-2 reasoning mostly (Nagel, 1995; Camerer, Ho, & Chong, 2004).

2.4 Anchoring and Adjustment

> We are merely reminding ourselves that human decisions affecting the future, whether personal or political or economic, cannot depend on strict mathematical expectation, since the basis for making such calculations does not exist; and that it is our innate urge to activity which makes the wheels go round, our rational selves choosing between the alternatives as best we are able, calculating where we can, but often falling back for our motive on whim or sentiment or chance. (Keynes, 1936, pp. 162- 163)

Depth of reasoning is a game theoretical model of reasoning influenced by the bounded rationality approach. Model implies while reasoning in the beauty contest game, individuals iteratively do mathematical calculations based on the self-generated or given reference point to reach the satisficing point. Dept of reasoning is a game theoretical model to understand competitors' behavior. There is also a behavioral economical approach which explains the individuals' judgement under uncertainty based on the given reference point: anchoring (Tversky and Kahneman, 1974).

Anchoring is a heuristic represented by Amos Tversky and Daniel Kahneman (1974) in the article "Judgement under Uncertainty: Heuristics and Biases". Kahneman explains the anchoring and adjustment with several examples such as the game of "wheel of fortune" or "multiplication". In the game of multiplication two groups of people asked to make the calculation of "1 x 2 x 3 x 4 x 5 x 6 x 7 x 8" and "8 x 7 x 6 x 5 x 4 x 3 x 2 x 1" in five seconds. The answer of the multiplication is "42,320". When it is assumed that the individuals are going to start the multiplication from left to right, the anchor is the multiplication which is already done in the inception of the mathematical process. Because of the "insufficient adjustment", the median guess of the individuals who are exposed to the series of the multiplication starting from one is "512" and the starting from eight is "2,250" (Tversky and Kahneman, 1974).

While the depth of reasoning can be observed in the guessing game, the research domain of the anchoring effect is broader. Anchoring heuristics is being studied for general knowledge/factual question, probability estimates, legal judgments, valuations/ purchasing decision, forecasting, negotiation, self-efficacy. There are three different mechanisms and types of anchoring suggested for different kinds of priming. Although there are objective mechanisms of the anchoring effect, there are studies shows that human factors such as mood, experience and cognitive ability also affects anchoring (Furnham and Boo, 2011).

Anchoring and adjustment mechanism are proposed by Tversky and Kahneman (1974). Mechanism suggests that anchoring is a conscious process of heuristics which uses the presented value or key as a reference of plausible answers. Suggested biases occurring insufficient adjustment or subjective probability distribution evaluation. Even though the suggested anchoring and adjustment mechanism is conscious and effortful, Kahneman and Tversky support that a non-informative priming affects the judgement under uncertainty as well. In the game of "wheel of fortune", players asked to answer two questions respectively after turning a wheel. The wheel is gimmick and only stops at "10" or "65". The questions are whether if the percentage of African nations in U.N. is larger than the number, the best guess of the player for the percentage. Since there is no relation between wheel of fortune with the factual question, no difference between "10" and "65" condition is expected. However, there is a significant difference between these two conditions is found. Median of the estimates are 25% for "10" and 45 for "65" (Tversky and Kahneman 1974).

Epley and Gilovich (2001), provided set of questions to the participants which activates internally self-generated anchors. Open ended questions are asked without any cue to anchor. Verbal reports of the participants about the strategies used revealed an anchoring heuristic underlying the answers. When the individuals asked to answer which degree does vodka freezes, it is reported from participants that the correct answer is not 0 Celsius since the common knowledge says that it is the freezing point of the water. The self-generated anchor is 0 Celsius and judgment is adjusted through lower degrees (*Mean* = -18.75 Celsius). There is a difference between externally provided anchors and self-generated anchors. Self-generated anchors are affected from incentives and forewarnings (Epley and Gilovich, 2005). It is also suggested there is a difference underlying the mechanism between self-generated and externally provided anchors. While self-generated anchors are known to be wrong and needs adjustment, anchoring and adjustment approach is accepted. However, externally provided anchors shows consisting confirmatory hypothesis testing (Epley and Gilovich, 2001).

Selective accessibility model takes anchoring as an activation to search for the relevant attributions (Strack and Mussweiler, 1997; Chapman and Johnson, 1999). Externally provided or self-generated anchors are taken into account as plausible answers. If the hypothesis that "the anchor is the true answer" is falsified, then a consisting and confirmatory search begins to reach an answer which has relevant attributes with anchor (Chapman and Johnson, 1994; Strack and Mussweiler, 1997). This model is reported as the dominant view in the anchoring framework (Furnham and Boo, 2011).

Wegener et al. (2001) suggest a model of anchoring based on the attitude change. While the models of anchoring and adjustment and selective accessibility proposes that the increase of the anchor extremity in plausible range also increases the anchoring effect when compared to plausible anchors. Attitude change model suggests that counter arguments occur when individuals are exposed the implausibly extreme values and causes smaller or none anchoring effect (Wegener et al., 2001).

Anchoring and adjustment and selective accessibility models are pointing the high-elaborative processes. Attitude change model handles low-elaborative process as well (Wegener et al, 2010). Blankenship et al. (2008) conduct a series of experiment using high cognitive load to reduce the elaboration on the judgment process. The cognitive load is increased by the multi-tasking during the question phase. It is seen that high elaborative process increases the usage of background knowledge, persistent over-time and less likely to affected by social influence. Also, it is reported that numerical priming has a lesser effect on anchoring while high elaborative process occurs. (Blankenship et. al., 2008, Wegener et al., 2010). In low elaboration process, anchors assumed to taken as a hint or cue to the correct answers (Wegener et al., 2010).

The camp of judgment and decision opposes the attitudes and persuasion camp by criticizing the elaboration likelihood model with “unnecessarily complex” duality of anchoring effect while appreciating the work. It is put that whether the input’s nature creates the additive effect or the traditional dual process system already elucidates the difference (Friedrick, Kahneman and Mochon, 2010).

It is expected in the present study that the gaze-cue sharing deviates the depth of reasoning and anchoring effect. Level-k model and anchoring effects occur in the uncertain environment. Gaze-cue sharing in this sense will prevent the uncertainty by broadcasting all participants states simultaneously. Given p-value makes the players guess either higher or lower than the rest of the players. Hence by following the cues of the other players, it may become easier to deduce reaching *NE*.

Multi-tasking increases the cognitive load (Blankenship et. al, 2008). In the gaze-cue sharing condition players are going to exposed to others player’s gazes as different colored cursors. Hence, not only the statistics given before the question but also cursors on the screen is going to be a concern for the players. Therefore, a reasoning process is going to be affected by tracking the other players. This additional task can cause higher cognitive load. It is expected the increase of the cognitive load has an effect on the judgment of the player (Blankenship et al. 2008; Wegener et al. 2010).

Instead of a single numerical information, current study presents a simultaneous flow of information related to the players future decision. Hence rather than a static anchor, players anchor question might be dynamic and change throughout the game. Stationary convergence of the cursors representing players on the specific numbers might adjust the anchor. This adjustment can deviate the judgment made by players.

CHAPTER 3

MATERIALS AND METHODS

3.1 Participants

There are 19 female (*Mean* = 25.684, *SD* = 4.726) and 11 male (*Mean* = 24.363, *SD* = 2.378) participants attended to experiment amassing 30 participants. There have been six groups of people taken as group of five people. Since there is a screen between computers and since the stimuli which represents other players do not give any clue about the participants' gender, groups have not been divided according to a gender category. All participants received average of 20 TL (3.29 US Dollars). In a session a group of five players is contested. While winner, as a player who wins the most trials, is rewarded with 30 TL, for each rank decreasing reward money is decreased 5 TL. Hence, the winner gets 30 TL, a player in the second place gets 25 TL, third player gets 20 TL and so on. Even a player wins no trials the base of 10 TL is given as the reward of the tailender player.

3.2 Stimuli

The experiment page consists of a ruler, "P" value and gaze cues. Ruler is set horizontally in the middle of the screen from the pixel number "209" to pixel number "834. The ruler shows only the decimal places with numbers and large notches. On the other hand, there is a notch for every single number on the ruler. P value is set 50 px above the center of the screen. There are three types of P values used in the experiment "1/2", "2/3" and "4/3" (Nagel, 1995).

There are three types of gaze cues used in the experiment as a stimuli: no, self, all gaze-cue sharing conditions. Gaze cues are the colorful ball-shaped indicators that indicate where a participant is looking on the ruler. Colors are specifically attributed to participants according to the order of connecting the server. The first participant who connects to the server is presented with a "red" (*color code:* **fa6161**) indicator, second participant is with a "green" (*color code:* **61fa61**) indicator and respectively others are presented with "blue" (*color code:* **6161fa**), "turquoise" (*color code:* **61fafa**) and "gray" (*color code:* **616161**). Although these colors are automatically assigned by Group Eye Tracking (GET) platform and reproduced for more participants, the colors stated above are the ones which are used in the experiments. *Figure 3.1* represents an experimental screen for "1/2" p value condition and all gaze-cue sharing condition. For other p values, number on the center of the screen changes as the condition changed. In "No Gaze Cue" condition there is no indicator of where other participants are looking. In "self gaze-Cue" condition there is only one indicator and the indicator on the ruler shows where the participant is looking on the ruler. However, in "All Gaze Cue" condition participants are seeing all indicators, in another saying, all participants are aware of where are the other participants are looking.

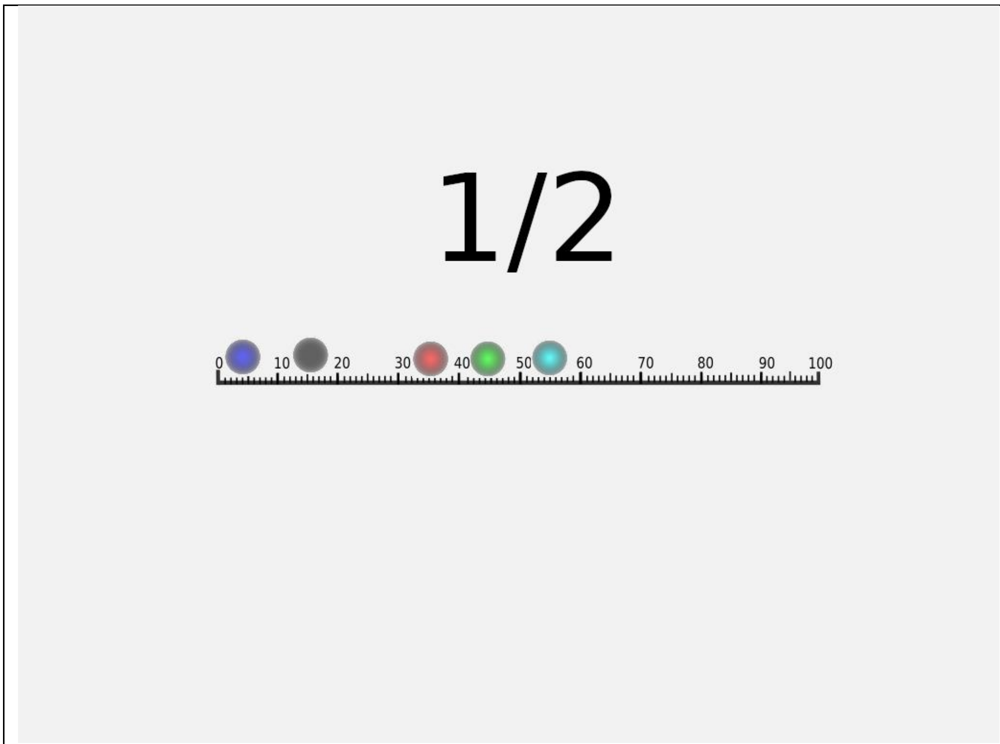
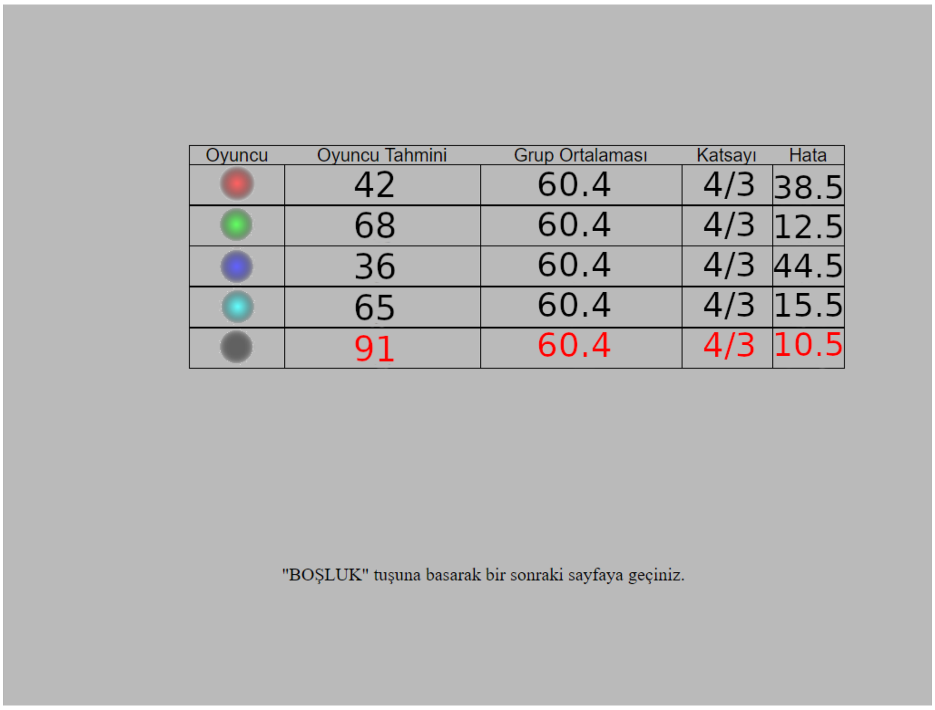


Figure 3.1: A trial for all gaze-cue sharing condition for p value of "1/2"

After each trial a score page is displayed to inform the participants. Score page is designed to inform the participants about the descriptive analysis and results of the trial in a broad manner. In the page all participants are represented as the color of their indicators under the player ("Oyuncu") column. For all rows, player’s guesses ("Oyuncu Tahmini"), average guess ("Grup Ortalamasi"), p - value ("Katsayi") and score ("Hata") is displayed in a table after all players finished the trial. Since the experiment is done in Turkish, while player guess and average guess is translated directly, p value is translated in the table as "constant" and score is translated as "error". The reason of translating score as error is to give the intuition that the lowest score represents lowest absolute distance of the guess and target number. Therefore, the intuition given is that the lowest score is the winner. Winner, as the lowest error, is painted in red. *Figure 3.2* is an example of score page used in the actual experiment. For instance, in Figure 3.2, the red player guessed 42, when the group average turned out to be 60.4. Since the p value was set at 4/3, the group members whose guess was closer to 60.4 x 4/3 = 80.53 was the winner. For the red player the error (score) was 80.53 - 42 = 38.5



Oyuncu	Oyuncu Tahmini	Grup Ortalaması	Katsayı	Hata
	42	60.4	4/3	38.5
	68	60.4	4/3	12.5
	36	60.4	4/3	44.5
	65	60.4	4/3	15.5
	91	60.4	4/3	10.5

Figure 3.2: A score page for p value of "4/3"

3.3 Design

The experiments were designed with two factors which have three discrete levels and nine conditions in total. The two factors are respectively the visual cue (gaze-cue sharing) a participant received while solving the problem and the different rates of p value. The three levels of gaze cue are: 1 - No cue given to any of participants, 2 - A participant only acquires their own gaze cue, 3 - All participants exposed to all participants gaze cue. The three levels of p value is defined by the values that has been used by Nagel in the original study: “1/2”, “2/3” and “4/3” (Nagel, 1995). Each condition is exposed to participants for four times repeatedly. This repetition unearths a learning effect between trials and can be taken as a factor as well. Experiment is designed as 3x3x4 repeated design hence all the conditions are exposed to all the participants.

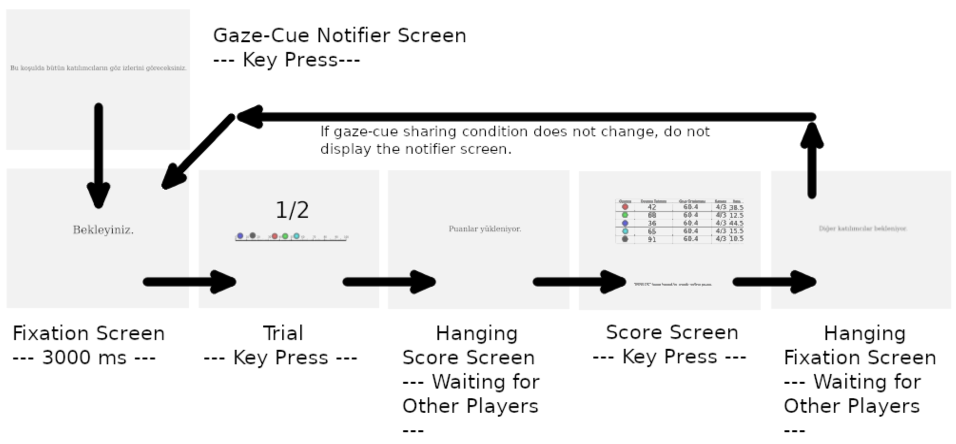


Figure 3.3: A simple workflow of the experiment

3.4 Group Eye Tracking Platform and Physical Environment

GET platform is designed by Ozan Deniz, Mehmetcan Fal and Cengiz Acartürk (Deniz, 2016; Deniz, Fal, & Acartürk., 2015) using Java language. The Platform is re-written in 2018 by Mani Tajaddini (Tajaddini, 2018) and Node.Js Javascript language is used. The experiment is written by the author [1] using the manual written by Mine Özkul (Özkul, 2018).

[1] With a significant help of Mine Özkul and Mani Tajaddini.

Figure 3.3 demonstrates the software architecture of the platform. "Main Server Code", "Base Experiment" and "Experiment Configurer" is a general script which is not written specific for this experiment. In general N clients are communicating with Main Server Code using Data collected by eye-tracker and keyboard. "Custom Experiment" is a script written by the author specific to the experiment. It provides tasks and user interface for the participant. Experiment specific message is given to the clients by "Custom Experiment" and status of the experiment to the base experiment module. Base experiment module is connected to the "Experiment Configuration and "Main Server Code". "Experiment Configurer" which holds the settings for "Base Experiment" can be adjusted by "Custom Configurer Page". Main Server Code publishes an automatic logging simultaneously in json object format.

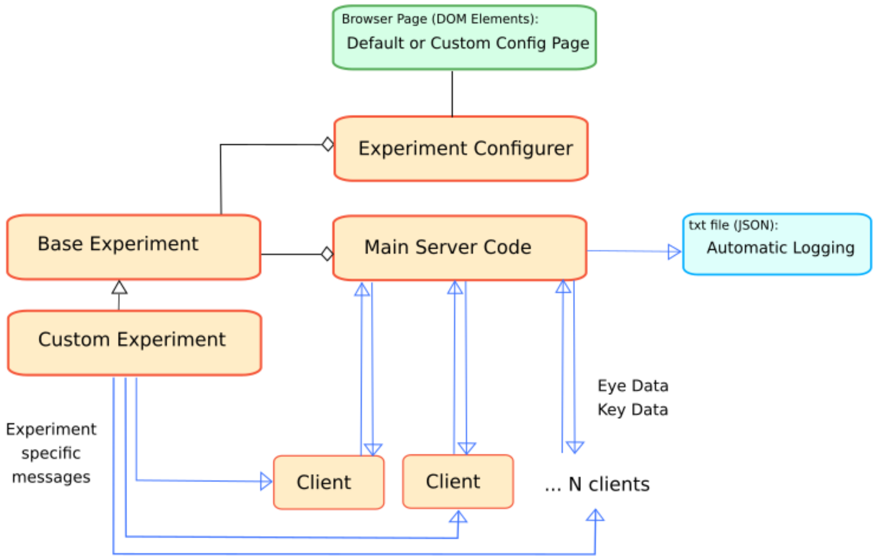


Figure 3.4: Basic workflow scheme of GET platform

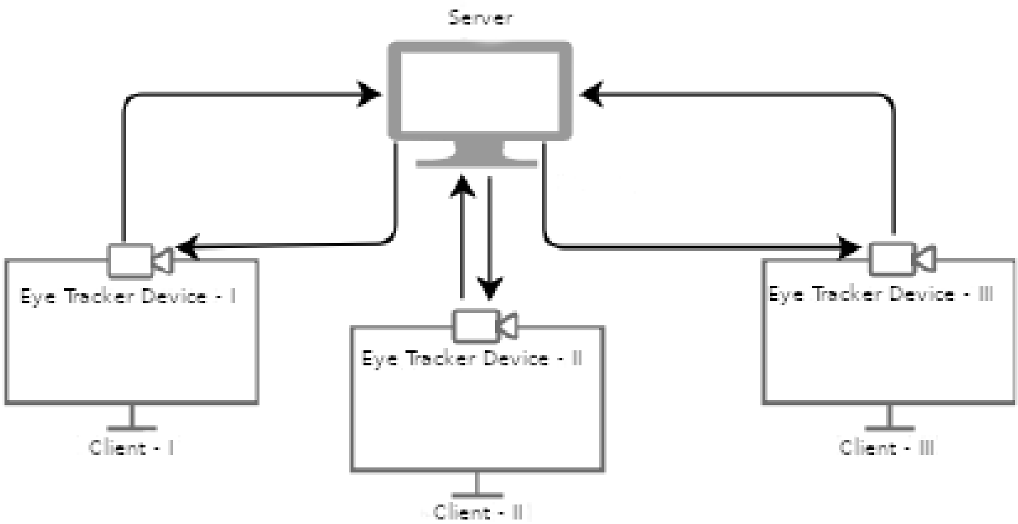


Figure 3.5: Basic workflow of the GET Hardware for three client. It must be noted that current study is conducted with five clients for all sessions.

EyeTribe is used as eye tracker device. "Main Server Code" is connected to EyeTribe API to collect the raw data. Sampling rate of the EyeTribe is configured to 60 Hz.. EyeTribes reported accuracy is 0.5 degrees and reported spatial resolution is 0.1 degrees. To smooth the gaze-cue sharing, average of the stack of ten samplings is used. Before each experiment EyeTribe's default nine-point calibration is supplied to each participant. Since the chin-rests (located exactly "54" centimeters (cm) away from screen) stabilize the head movements, there is no other calibration made for the experiment. No calibration deterioration occasion during experiment is reported by participants.

3.5 Procedure and Data Preparation

This section covers the procedures followed by the researchers while conducting the experiment and the preparation of the data such as detecting the eye-movements.

3.5.1 Procedure

The Experiments was conducted in METU Informatics Institute Laboratory. GET platform and server were located in the same laboratory. The laboratory has six computers. Five of them are the computers that experiment was conducted and one of them is server computer. There are panels between computers to prevent the interaction between participants. Even though panel does not exactly block people from seeing each other, it blocks seeing other participants' screen. There is also a chin rest in front of every computer which experiment carried out. Chin rest is used to stabilize participants' head as make them not broke the calibration.

Figure 3.6: METU Informatics Institute Group Eye Tracking Laboratory

Participants were taken in the room five by five. There was always only one researcher in the room with participants during the experiment. After all participants arrived, consent form was given to them with a demographic information paper (See Appendix A for consent form and Appendix B for demographic information paper). Then, an instructions paper was given (See Appendix C for instructions paper). In the given and read aloud instructions paper what participants' have to follow was written. Before the experiment researcher asked whether all participants read the instructions paper. If all of the participants read the paper, then they were allowed to ask questions about the game. While researchers were answering the questions, they meticulously paid attention never to use any number to misdirect the participants or create a bias.

Figure 3.7: Author testing the exercise documented by two different angles

After the question part is over, researcher adjusted the chin rest according to participants' height and calibrated the eye tracker according to participants' eyes. If eye-tracker software was not given the perfect score, then calibration test has been redone with a guidance of a researcher. After all participants' chin rest adjustment and eye-tracker calibration were made then the exercise part was started.

Exercise part was designed as a three question with all the gaze cue conditions respectively. However, P value set to "1" all the time. To reduce misdirection, researches warned participants three times about the P value. While the exercise part was conducting, researchers asked participants about whether they are comfortable with their chin rest and eye tracker calibration. If participants were not comfortable, then adjustments and calibration were redone until the participant's validation.

Experiment was conducted after participants read the instructions paper, chin rest was adjusted, eye tracker is calibrated and exercise part was done. After the experiment is started, there is no talking or asking questions allowed.

Guessing procedure is divided into two phases. The first phase is called the "choosing phase". Participants are expected to choose an exact point in the ruler by looking at it and press “SPACE”. Anyone who presses space then pass into second phase which is called "decision phase". In decision phase the indicator gets frozen and participant cannot see other participants. Meanwhile, other participants cannot see the participant in the decision phase. Since the indicator is freezes, eye movements can no longer moves the eye tracker and so it cannot change the decision made. In this phase it is expected from participants to adjust their guesses. A question "Seçiminiz?" (Your choice?), appears above the indicator near the quantitative equivalent of the place that have been chosen on the ruler. In the decision phase, participant has a chance to increase or decrease their guess for five points. While increasing or decreasing their guesses they have to use "LEFT" key to decrease the guessed value or "RIGHT" key to increase the guessed value. If a participant wants to finalize the decision made she has to press "DOWN" button. Any participant who finalizes their decision is going to be sent to the score screen.

Guessing game is played for 36 times with no break. After every single game, the score of the game is shown. Experiment is finished with a final score screen. Final score screen includes which player wins how many games. Researcher allocates the reward money according to winning times. Mean of the reward gained by participants is 20 TL. In the group of five, the participants that wins the most is rewarded with 30 TL, the second is rewarded with 25 TL and it goes with 20 TL, 15 TL and 10 TL respectively. If two players win the same amount of games, they are rewarded with the mean of their standings. For example, if the third and the fourth participants' win amount is same, they both gain 17.5 TL. If second, third and the fourth participants win amount is same, they are both awarded with 20 TL.

There are three types of data are collected with GET platform: "gaze log", "keyboard log" and "score log". Gaze Data is being logged in 60 Hz for every participants' gaze on the screen as an X and Y coordinates. It also collects the id, server timestamp, client timestamp, difference between server and client time stamp, "P Value", "Gaze Condition", guessed number, group size (5 for all cases), screen type and question number. Keyboard Data consists of id, server time stamp, client time stamp, time difference, key pressed, group size (5 for all cases), screen type and question no. Lastly, score log consists of score of the all participants and server timestamp which is logged after the final participant made her decision. Data preparation is made using *pandas* library (McKinney, 2010) in Python programming language.

For analysis, raw gaze data is used. To test the guesses, for each trial and each participant, latest guess value recorded for each player is extracted. For the guess times, for each participant, the duration of the experiment as between the first data recorded in the experiment screen and the last data in the experiment per question is extracted. All of these data are written as individual guess log and individual guess time log. For each question a unique code is generated based on the combination of the conditions featured in question such as "1-1/2-ALL". This code means that the question is the first trial of the all gaze-cue sharing condition for p-value of "1/2". Hence there is no question for player to play exactly this combination of the conditions more than once. Each column extracted from a participant's gaze log is manually merged matching their combination of conditions.

For guess values the final and adjusted decision of the participant is used. For each gaze-cue sharing condition, when a participant presses the "SPACE" key, screen freezes with an indicator written "Seçiminiz?" ("Your choice?") above the gaze-cue. When a player presses "SPACE" player's gaze-cue indicator vanishes from other players' screen and other players' gaze-cue indicator vanishes from the player's screen. Also the number guessed is written numerically. In that point, the final adjustment can be made with the distance of five number from the already chosen number. There are two main reasons to choose the final adjustments rather than the initial guesses:

- The sampling rate of an eye tracker is 60 Hz.. Yet there can be smooth pursuits or saccade cannot be detected just before or simultaneous with pressing the "SPACE" key. A player should have a right to adjust her answers within a range to avoid losing a motivation against game and trust against experiment. However, as not expected, no participant ever reported the guessing errors because of the devices or GET platform.
- Since players' gaze-cue indicators, in other words possible guesses and guess tendencies, can be observed by other players, it is a strategy to vanish before the actual guess. Hence player's true motivation and decision can be read by detecting the adjusted guesses rather than initial guesses.

The current state of the GET does not classify the eye movement itself. The eye movements are classified using I-VT filter. The I-VT filter separates eye movements by point – to – point velocities. Velocity below 100 deg/sec identifies as fixation and above 300 deg/sec identifies as saccade. Between two is classified as mixed (Salvucci & Goldberg, 2000). The filter is provided to the individual gaze log separated before. To use the filter, height of the screen (30 cm), the distance of the players away from the screen (56 cm) and vertical resolution of the screen resolution of the (768 px) is used. The code is given in the Appendix D.

Code for the fixation duration calculation the difference server time stamp between of a fixation start after a "saccade" or "mixed" eye movement and ends before "mixed" or "saccade" eye movement. And average of the fixation duration is calculated. Average saccadic amplitude is calculated using the x and y coordination before saccadic eye movement and after saccadic eye movement. In this sense, the distance when a saccadic eye movement occur is calculated. The average of saccadic amplitude is generated by diving this total distance of amplitude with amount of saccade observed. Saccade frequency is calculated as detected saccade per second. Frequency is used rather than count to prevent the confounding variable of guess times.

CHAPTER 4

RESULTS

The games are played with 5 participants in each group for three different conditions. The three conditions are "no gaze-cue", "self gaze-cue" and "all gaze-cue". The rule of the game is to choose a number between 0 and 100 and the key to win is to choose the number which is closer to the target number. The target number is calculated with the multiplication of the "p" constant and the mean of all guesses in the round. There three p values defined in the game; "1/2", "2/3" and "4/3" (Nagel, 1995). For each conditions the game is played with four identical rounds. Analyses are made with 3x3x4 repeated ANOVA test for gaze-cue sharing conditions, p values and trials. 3x4 repeated ANOVA is used to test the "guess values" since it is expected that p-values higher than and less "1", by its nature, acquire different directions of guess values to win the game. The direction regulation is made using the distance from equilibrium. Since the equilibrium is "0" for "1/2" and "2/3", the distance between equilibrium is the exact number of the guess itself. For the "4/3" equilibrium is "100". Therefore, the distance from equilibrium is the absolute value of "100" minus guess. Analyses of the guess values are actually the analysis of the guesses' distance from the equilibrium.

Guess Time analysis is made at the level of milliseconds. Greenhouse-Geisser sphericity correction test is used, for 3X4 repeated design. Guessing procedure consists of two phases. The first phase is "guessing" phase which takes between exposition of the trial and making a coarse decision. The coarse decision is made with the player looking for the guessed number on a ruler on the screen and pressing "SPACE" key. Second phase takes time between the exposition of the trial and making an ultimate decision. After making a coarse decision, players' screen freezes. To make an ultimate decision for trial, players' have a right to change their decision in an absolute distance of five by pressing "LEFT" and "RIGHT" keys. After having a definite guess, players ultimate their guess by pressing "DOWN" key. Thus, guess times are measured as the times between exposition of the trial and pressing "DOWN" key.

4.1 Results for Guess Values

Guess values are regulated as the distance from the equilibrium. Distance from the equilibrium is calculated as:

$$Distance(guess, p) = \begin{cases} |0 - guess|, & p < 1 \\ |100 - guess|, & p \geq 1 \end{cases}$$

According to this formula, 4/3 guesses are regulated as the distance from "100". For the descriptive plots in *Figure 4.1, Figure 4.2 and Figure 4.3*, decreasing trends are seen form first trials to the last trials. However, there is no clear pattern in the gaze-cue sharing conditions. For both "2/3" and "4/3" the decreasing trend pattern is violated from second to third trial. Another interesting point is seen in the "1/2" gaze p value's none gaze-cue condition. It is seen that trend is low but after a while it slightly increases.

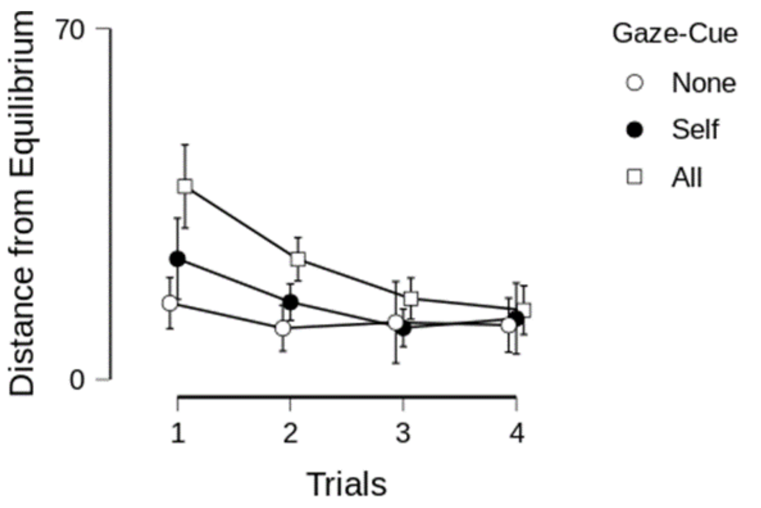


Figure 4.1: Descriptive plot for *"p = 1/2"* distance of guesses from equilibrium distinguished by gaze-cue conditions for four trials

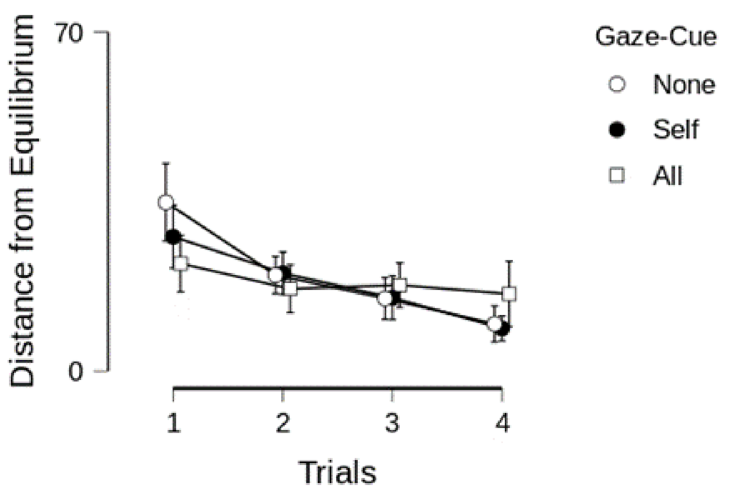


Figure 4.2: Descriptive plot for *"p = 2/3"* distance of guesses from equilibrium distinguished by gaze-cue conditions for four trials

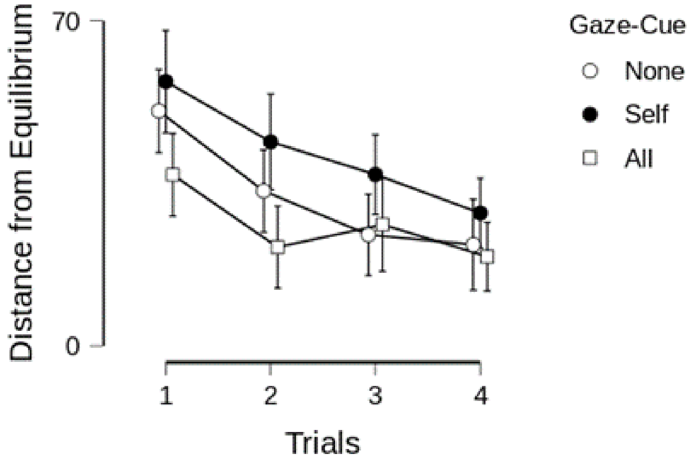


Figure 4.3: Descriptive plot for *"p = 4/3"* distance of guesses from equilibrium distinguished by gaze-cue conditions for four trials

The distance from equilibrium analysis is made with 3x3x4 repeated ANOVA. Sphericity is violated and Greenhouse-Geiser correction results are used. As shown in Table E4.1, there is no difference between guesses' distance from equilibrium ($F(1.640, 47.548) = 1.967$, $p = 0.158$, *partial* $\eta^2 = 0.064$). The p value effect for the distance is significant ($F(1.129, 32.720) = 2.001$, $p < 0.001$, *partial* $\eta^2 = 0.431$). Also, trials are differed significantly ($F(2.383, 69.105) = 69.641$, $p < 0.001$, *partial* $\eta^2 = 0.706$). The interaction effect of gaze-cue and p value ($F(1.898, 55.039) = 7.825$, $p = 0.001$, *partial* $\eta^2 = 0.212$) and p value and trial ($F(3.920, 113.676) = 2.930$, $p = 0.025$, *partial* $\eta^2 = 0.092$) are significant. While there is no significant interaction is seen in gaze cue sharing conditions and trials ($F(3.949, 114.513) = 0.975$, $p = 0.423$, *partial* $\eta^2 = 0.033$), when all factors is taken together there is significant interaction observed ($F(6.413, 185.968) = 3.908$, $p < 0.001$, *partial* $\eta^2 = 0.119$).

Follow up Post-Hoc analysis does not show a difference between "1/2" and "2/3" p values conditions (*Mean Difference* = -1.856, $t = -0.683$, $p_{holm} = 0.497$). However "1/2" (*Mean Difference* = -16.447, $t = -6.056$, $p_{holm} < 0.001$) and "2/3" (*Mean Difference* = -14.592, $t = -5.372$, $p_{holm} < 0.001$) is significantly different when compared to "4/3" p value (See Table E4.2 for the Post-Hoc comparisons of p-values).

Convergence to equilibrium can be seen from the post-hoc analysis of the trials (Table E4.3). The distance from guesses are significantly different between trial 1 and the rest: trial 2 (*Mean Difference* = 11.296, t = 8.341, p_{holm} < 0.001), trial 3 (*Mean Difference* = 14.911, t = 11.010, p_{holm} < 0.001) and trial 4 (*Mean Difference* = 18.437, t = 13.614, p_{holm} < 0.001). Difference between trial 2 and trial 3 is confusing in the descriptive plot, however analysis shows a significant difference between trial 2 and trial 3 (*Mean Difference* = 3.615, t = 2.669, p_{holm} = 0.018). Also, trial 2 and trial 4 is significantly different as well (*Mean Difference* = 7.141, t = 5.273, p_{holm} < 0.001). Finally, the difference between trial 3 and trial 4 is also significant (*Mean Difference* = 3.526, t = 2.604, p_{holm} < 0.018).

4.2 Results for Guess Times

Figure 4.4, *Figure 4.5 and Figure 4.6* display plots for guess times for 1/2 of p-value, 2/3 of p-value and 4/3 of p-value, respectively. Guess times are the times takes between the exposition of the trial and the final decision. Each player has the right to set her decision five points up or down using the keyboard. Pressing "DOWN" key finalize the adjustment. Final decision is not the time that a participant presses "SPACE" key on the keyboard to finish the eye-track on the ruler, but the time that she/he presses "DOWN" after making final adjustments. Three figures demonstrate guess times by p value in the separated lines of gaze-cue sharing conditions by trials. It is seen that none and self gaze-cue conditions have similar guess time values and decreasing trends from first to fourth trial. Nevertheless, all gaze-cue sharing conditions have higher guess times and different trends towards the fourth trial and different gaze-cue. In general, it can be said that there is a decreasing trend towards the later trials, however, *p = "1/2"* has a highest value for second trial and *p = "4/3"* has a higher value of the fourth trial than third trial. It must also be noted that while none and self gaze-cue sharing conditions has similar deviation bars, all gaze-cue sharing condition has a decreasing standard deviation bars between towards the last trial in general.

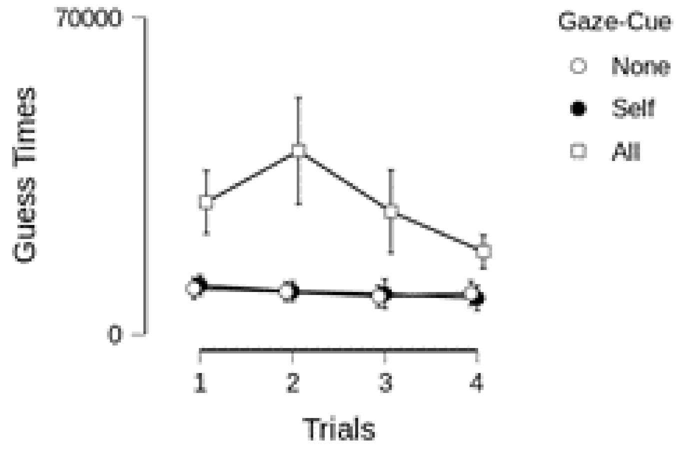


Figure 4.4: Descriptive plot for *"p = 1/2"* guess time distinguished by gaze-cue condition for four trials

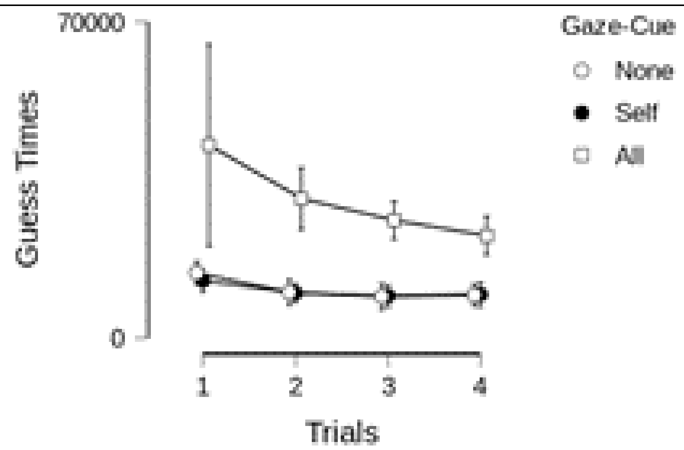


Figure 4.5: Descriptive plot for *"p = 2/3"* guess time distinguished by gaze-cue condition for four trials

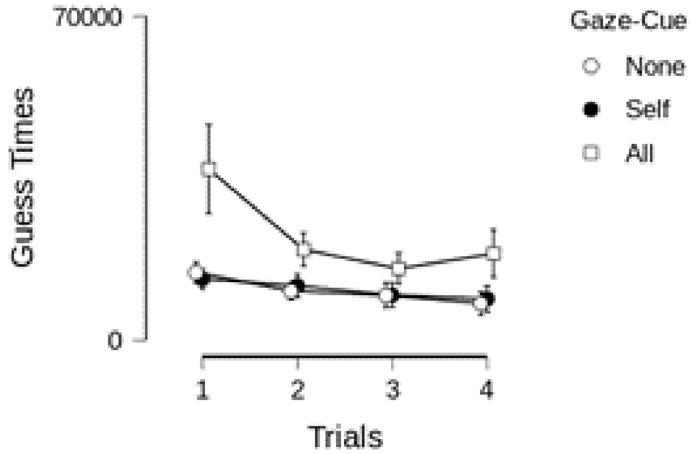


Figure 4.6: Descriptive plot for *"p = 4/3"* guess time distinguished by gaze-cue condition for four trials

3x3x4 repeated ANOVA is used to test the differentiation of the guess times. It is found that the different categories of gaze-cue sharing condition makes different guess times ($F(1.058, 30.684) = 45.093$, $p < 0.001$, *partial* $\eta^2 = 0.609$). As it can be seen from the Table E4, there is a significant difference for trials as well ($F(1.955, 56.699) = 3.033$, $p < 0.001$, *partial* $\eta^2 = 0.269$) but no difference for p - values ($F(1.955, 56.699) = 3.033$, $p = 0.057$, *partial* $\eta^2 = 0.095$). Gaze-cue sharing conditions have significant interaction with p values ($F(2.111, 61.227) = 4.338$, $p = 0.016$, *partial* $\eta^2 = 0.130$) and trials ($F(1.566, 45.407) = 4.170$, $p = 0.030$, *partial* $\eta^2 = 0.126$). As well, p value and trials has a significant interaction as well ($F(2.507, 72.710) = 3.255$, $p = 0.034$, *partial* $\eta^2 = 0.101$). All in all, there is no interaction when all these three factors come together ($F(2.333, 67.662) = 2.666$, $p = 0.068$, *partial* $\eta^2 = 0.084$).

Holm's correction post-hoc test is used to further investigation and between gaze-cue condition (see Table E4.5 for post-hoc comparisons of guess times for gaze-cue condition). There is significant difference between all gaze-cue sharing condition and the rest: none gaze sharing condition (*Mean Difference* = 17066.558, $t = 8.237$, $p_{holm} < 0.001$) and self gaze-cue sharing condition (*Mean Difference* = 17121.403, $t = 8.237$, $p_{holm} < 0.001$). In terms of guess times, no significant difference is found between none gaze-cue sharing and self gaze-cue sharing conditions (*Mean Difference* = 54.844, $t = 0.026$, $p_{holm} = 0.979$).

There has been four trials and a significant difference of guess times between these trials lead to a post-hoc test of Holm's correction (see Table E4.6 for Post-Hoc comparisons of guess times for trials). Results demonstrate that the first trial makes a significantly slower guesses than third trial (*Mean Difference* = 6759.585, t = 4.336, p_{holm} < 0.001) and fourth trial (*Mean Difference* = 8011.830, t = 5.174, p_{holm} < 0.001). The same quickness is not observed for the second trial (*Mean Difference* = 3614.996, t = 2.335, p_{holm} = 0.066). Second trial shows no difference with third trial (*Mean Difference* = 3144.589, t = 2.031, p_{holm} = 0.091) but shows a difference with fourth trial (*Mean Difference* = 4396.833, t = 2.840, p_{holm} = 0.022). Lastly, there is no difference in guess times between third trial and fourth trial (*Mean Difference* = 1252.244, t = 0.089, p_{holm} = 0.421).

4.3 Results for Average Fixation Duration

The descriptive plots for average fixation duration for p-values for 1/2, 2/3 and 4/3 are given in the *Figure 4.7, Figure 4.8* and *Figure 4.9* respectively. The figure includes three plots separated by the p values given to the participants. Fixation duration refers to a time when a participant's gaze is locked to an approximate exact spot on the screen. To calculate the duration of the fixation, eye movements are classified by the velocity of participant's eye movement. Fixations are calculated by separating continuous fixations from "mixed" and "saccade" movements and having an average point from the stack. It is seen from the descriptive plot that fixation duration values for gaze-cue sharing conditions, p values and trials overlaps.

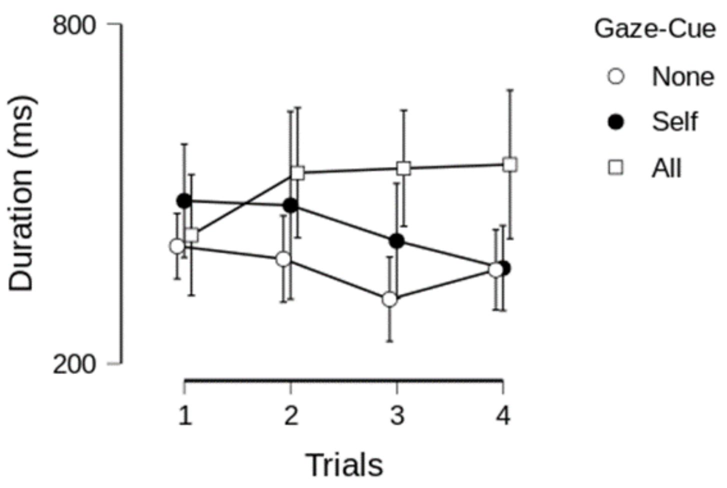


Figure 4.7: Descriptive plot for *"p = 1/2"* fixation duration distinguished by gaze-cue conditions for four trials

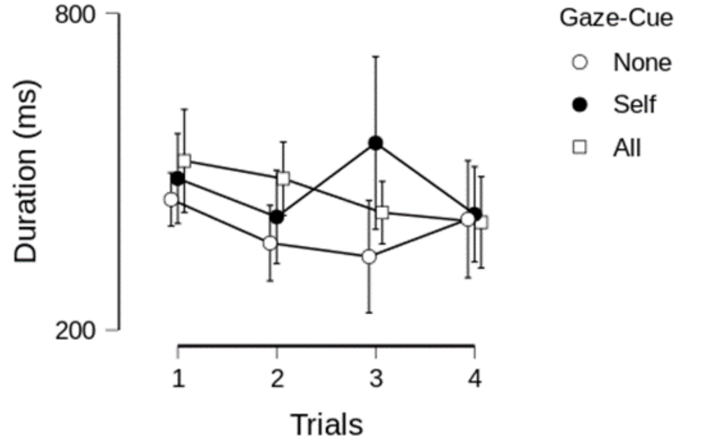


Figure 4.8: Descriptive plot for *"p = 2/3"* fixation duration distinguished by gaze-cue conditions for four trials

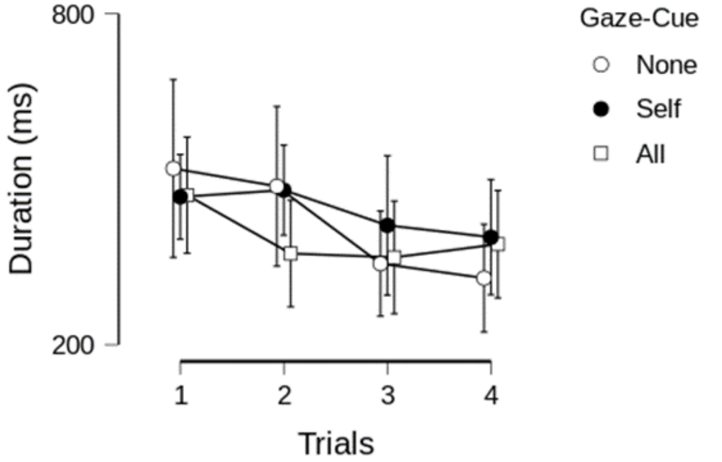


Figure 4.9: Descriptive plot for *"p = 4/3"* fixation duration distinguished by gaze-cue conditions for four trials

Gaze-cue sharing conditions in three levels (none, self and all gaze-cue sharing), p values in three levels (1/2, 2/3, 4/3) and trials in four levels (1 , 2, 3, 4) are analyzed with 3x3x4 repeated ANOVA measures. As seen in Table E4.7, there is no significant difference for the levels of gaze-cue sharing conditions ($F(1.729, 50.145) = 2.200$, $p = 0.128$, *partial* $\eta^2 = 0.071$), p values ($F(1.993, 57.748) = 0.762$, $p = 0.471$, *partial* $\eta^2 = 0.026$). There is a significant difference between trials ($F(1.842, 53.595) = 3.517$, $p = 0.040$, *partial* $\eta^2 = 0.108$). Interaction effect between gaze-cue sharing conditions and p values ($F(3.319, 96.241) = 2.891$, $p = 0.034$, *partial* $\eta^2 = 0.091$) is significant. Gaze-cue sharing and trials ($F(4.527, 131.276) = 1.118$, $p = 0.353$, *partial* $\eta^2 = 0.037$) is not significantly different. P-Value and Trials interaction also show no significant result ($F(4.562, 132.312) = 1.618$, $p = 0.165$, *partial* $\eta^2 = 0.053$). Together, interaction effect of gaze-cue sharing, P-value and trials also shows no significant outcome ($F(6.473, 187.720) = 1.448$, $p = 0.194$, *partial* $\eta^2 = 0.048$).

Follow up test for trials is made using Holm's correction. It is seen that the only difference between trials for fixation duration is between first and the last trial (*Mean Difference* = 68.661, $t = 2.916$, $p_{holm} = 0.027$). There is no difference between other trials given in the *Table E4.8*.

4.4 Results for Saccade Frequency

Saccade is ballistic eye movement in general more than 300 deg/sec (Salvucci & Goldberg, 2000). *Figure 4.10, Figure 4.11* and *Figure 4.12* show the plots comparing saccade frequency for gaze-cue sharing conditions in four trials for p values of 1/2, 2/3 and 4/3 respectively.

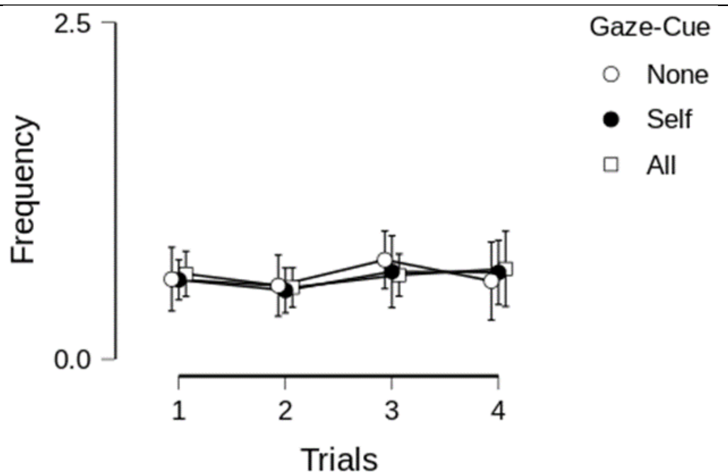


Figure 4.10: Descriptive plot for *"p = 1/2"* saccade frequency distinguished by gaze-cue conditions for four trials

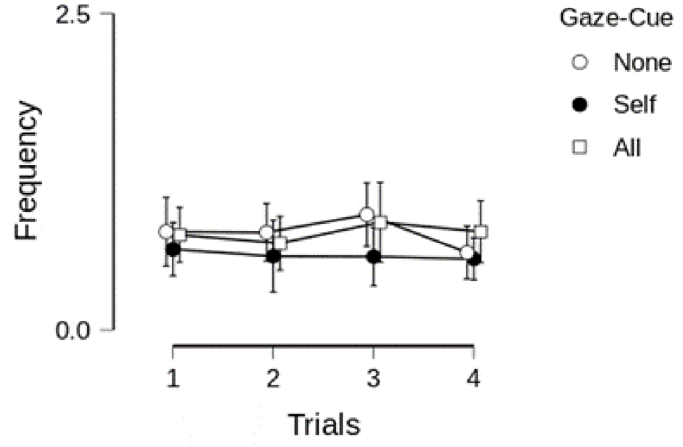


Figure 4.11: Descriptive plot for *"p = 2/3"* saccade frequency distinguished by gaze-cue conditions for four trials

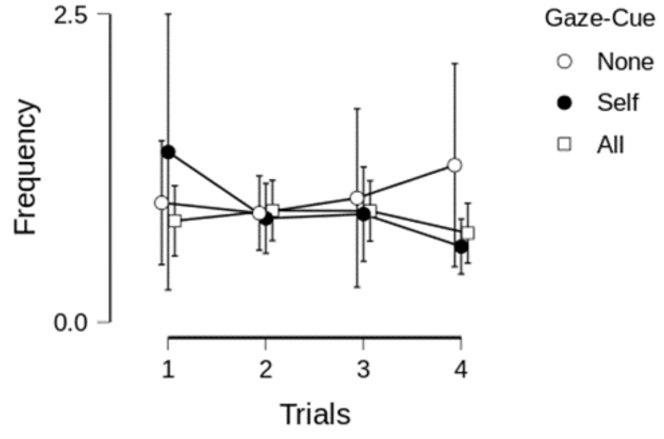


Figure 4.12: Descriptive plot for *"p = 4/3"* saccade frequency distinguished by gaze-cue conditions for four trials

To differences within subjects in terms of saccade frequency per second in levels of gaze-cue sharing, p value and trials 3x3x4 repeated ANOVA is used with Sphericity Correction (see Table E4.9). The outcome indicates that there is no deviation between levels of gaze-cue sharing ($F(1.622, 47.041) = 0.622$, $p = 0.509$, *partial* $\eta^2 = 0.021$), p value ($F(1.119, 32.447) = 2.136$, $p = 0.152$, *partial* $\eta^2 = 0.069$) and trials ($F(2.221, 64.416) = 1.271$, $p = 0.289$, *partial* $\eta^2 = 0.042$). Gaze-cue sharing and p value ($F(1.546, 44.832) = 0.628$, $p = 0.499$, *partial* $\eta^2 = 0.021$), gaze-cue sharing and trials ($F(2.910, 84.402) = 0.789$, $p = 0.500$, *partial* $\eta^2 = 0.026$), p value and trials ($F(3.483, 101.002) = 0.598$, $p = 0.642$, *partial* $\eta^2 = 0.020$) show no significant interaction. When all factors come together, there occurs no interaction effect as well ($F(3.019, 87.562) = 1.335$, $p = 0.268$, *partial* $\eta^2 = 0.044$).

4.5 Results for Saccadic Amplitude

Saccade amplitude is measured by calculating the distance between the former and the current point recorded when a saccade is detected in degrees (Salvucci & Goldberg, 2000). *Figure 4.13*, *Figure 4.14* and *Figure 4.15* display the descriptive plots for saccade amplitudes for p values of 1/2, 2/3 and 4/3 respectively. The horizontal axis of the plot refers to trials from 1 to 4 and the separate lines compares gaze-cue sharing conditions.

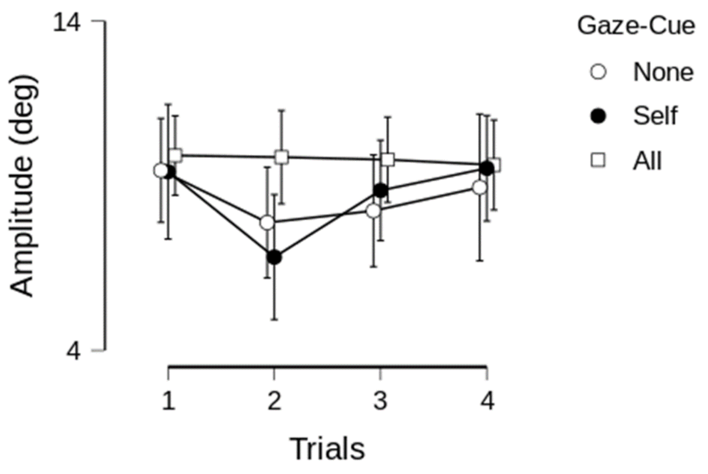


Figure 4.13: Descriptive plot for *"p = 1/2"* saccade amplitude distinguished by gaze-cue conditions for four trials

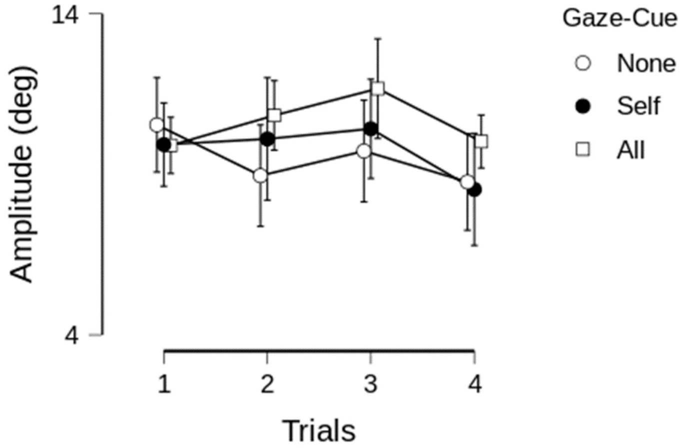


Figure 4.14: Descriptive plot for *"p = 2/3"* saccade amplitude distinguished by gaze-cue conditions for four trials

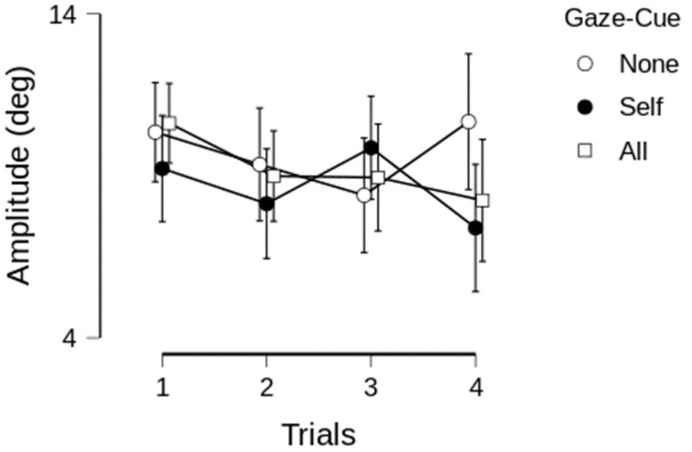


Figure 4.15: Descriptive plot for *"p = 4/3"* saccade amplitude distinguished by gaze-cue conditions for four trials

3x3x4 repeated ANOVA with Greenhouse-Geisser sphericity correction is made to test the differences of saccade amplitude between gaze-cue sharing condition, p values and trials A4.10. The outcome indicates saccade amplitudes do not significantly differ for gaze-cue sharing conditions ($F(1.843, 53.456) = 1.815$, $p = 0.175$, *partial* $\eta^2 = 0.059$), however they differ significantly for p values ($F(1.631, 47.296) = 4.024$, $p = 0.032$, *partial* $\eta^2 = 0.122$) and trials ($F(2.706, 78.473) = 3.224$, $p = 0.031$, *partial* $\eta^2 = 0.100$). For all the interaction combinations, the outcome of the test displays no significant interaction effect; gaze-cue and p value ($F(3.465, 100.477) = 1.681$, $p = 0.168$, *partial* $\eta^2 = 0.055$), gaze-cue and trials ($F(4.300, 124.707) = 1.268$, $p = 0.285$, *partial* $\eta^2 = 0.042$), p value and trials ($F(5.001, 145.034) = 1.527$, $p = 0.185$, *partial* $\eta^2 = 0.050$) and all together gaze-cue, p values and trials ($F(7.486, 217.091) = 1.106$, $p = 0.360$, *partial* $\eta^2 = 0.037$).

As shown in Table E4.11, Holm's correction post-hoc test reveals that saccadic amplitude 1/2 p value is significantly lower than 2/3 p-value (*Mean Difference* = -0.915, $t = -2.645$, $p_{holm} = 0.031$). No significant difference is found between 1/2 and 4/3 (*Mean Difference* = 0.756, $t = -0.434$, $p_{holm} = 0.666$) and 2/3 and 4/3 (*Mean Difference* = 0.765, $t = 2.221$, $p_{holm} = 0.062$).

The differences in saccadic amplitudes are seen between the trials. It is measured that saccade amplitude is significantly higher in first trial than second trial (*Mean Difference* = 0.934, $t = 2.730$, $p_{holm} = 0.046$). There is no significant difference between other trials. For post-hoc of saccade amplitude for trials, see *Table E4.12*.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Discussion of Guess Reults

For the comparison of the guess values, guesses are adjusted according to their distance from the equilibrium. Equilibrium of p values of "*1/2*" and "*2/3*" is going to hold the guesses converging to “0”, "*4/3*" is going to hold guesses converging to "100". It is considerable that this reverse directional convergence can cause a significant difference naturally if not adjusted. Results of the guesses values adjusted by the distance from the equilibrium is given in the *Table E4.1*. H_1 is falsified with results showing no difference between gaze-cue sharing conditions. Nagel (1995) found that convergence rate to equilibrium for the median is higher in "1/2" than "2/3". The present study on the other hand makes does not show any difference between gaze-cue sharing conditions. The reason of the indifference can be the difficulty of calculating p - values of the average.

Insufficiency of the provided certainty might be one of the reasons of the guess results. Although it is known previous and present behavior of the other players, the future is still in question. Tversky and Kahneman (1974) refer to the uncertain such as outcome of an election or future value of the dollar. Judgments to guess certain numbers or words in an uncertain environment is the spine of the anchoring studies. In general, these studies focus on the trivial questions rather than game theoretical competitions (Furnham and Boo, 2011). Hence the time series of an anchoring study goes as collecting background knowledge (t_{before}), exposing to an externally provided anchor or a question without an anchor ($t_{exposition}$), adjusting the possible answer, conducting a consisting hypothesis testing or constructing an attitude towards externally provided or self generated anchor ($t_{reasoning}$) and the end of the experiment results (t_{after}). While no and self gaze-cue sharing holds the same structure of time series with uncertainty elements, all gaze-cue sharing condition produce its own uncertainty. “$t_{reasoning}$” stage, besides holding the uncertainty of the t_{after}, holds the uncertainty of the outcome ($t_{reasoning+1}$) of the micro-decisions to make a strategic interaction with opponents. Every single step to $t_{reasoning+n}$ holds the uncertainty of the outcome of the players process of the planned action: going left, going right or staying.

While the prevention of the uncertainty is never mentioned in the study, the decrease of the uncertainty argument is supported with simultaneous information broadcasting to each player in all gaze-cue sharing condition. Watching other players while having a reasoning process on the guess might have increased the cognitive load. However, Blakenship et al. (2008) studied the cognitive load effect with multi-tasking. Three sceptic criticism are there to make to the present study's expected cognitive load suggestion. First criticism address' to the unknown motivation of the participants. A participant may or may not care for the given gaze-cues. There is no deductive way of detecting whether gaze-cues given were incentives for participants fit their decisions. Second criticism refers to the nature of the multi-tasking. Reasoning on a guess while following cursors on the screen is a multi-tasking if both of the moves' performance are measuring without a relation. If there is a correlation between the performance of the simultaneously executed behaviors of the individual, then it can be called as coordination as well. While it is possible to say that changing a task might have an effect on the cognitive load, assuming coordination cause the effect can be counterintuitive. Lastly, high elaboration and low elaboration thinking refers to thoughtful and non-thoughtful process. It is assumed that high cognitive load push competitors to make non-thoughtful process (Blakenship et al., 2008, Wegener et al., 2010). The controversy in here is to assume all gaze-cue condition increases cognitive load by dramatically increasing the amount of stimuli and attention needed causes low elaboration thinking a non-thoughtful process of judgment. The dynamic structure of the all gaze-cue sharing condition effect even if it can increase the cognitive load, does not have to decrease the elaboration of thinking. Hence even though deviation may occur by the dynamic structure and cognitive load, it does not show that it comes from a non-thoughtful process. Elaboration can also be related with individuals' cognitive abilities. However, literature is controversial for the cognitive ability and anchoring effect (Oechssler et al., 2009, Bergman et al., 2010).

For all p values, it is found that trials are significantly differentiated. Post-Hoc tests demonstrate that even though later trials gets closer, the first trials are always significantly less converged to equilibrium. Therefore, learning effect can be mentioned. When the interaction of the gaze-cue sharing conditions and trials are taken into account, there is no dependence found at *Table E4.1*. Learning effect across trials might be suppressed the effect of gaze-cue sharing. For every trial, p value condition and gaze-cue condition throughout the game, experience of the players to the condition increases. Experienced players are affected less from the anchoring (Chapman and Johnson 1994; Wilson et al., 1996). Repeated design can be criticized in this case. Learning after a while may create a habitual and recognized strategy or decision fatigue. Order of the exposure to gaze-cue sharing is counterbalanced for each session. A non-repeated and one-shot study can be designed to reduce the confounding factors.

5.2 Discussion of Guess Times Results

Guess times are tested with 3x3x4 repeated ANOVA. Again, with the "guess" analysis, "guess times" analysis is also made after the final adjustment. The reason is that the final adjustment gives the actual outcome of the game. Nevertheless, in this analyze (and the rest) p value is not separated because the difference in task has no mechanical difference for the outcome of what is expected. However, three different descriptive plots are presented in the *Figure 4.4, 4.5* and *4.6*. For all p values, a dramatic difference between all gaze-cue sharing and the rest is demonstrated.

H_2 is not falsified since there is a significant difference within gaze-cue sharing conditions. Mean difference between conditions revealed that for all trials and p values, all gaze-cue sharing conditions are significantly higher than the rest of the gaze-cue sharing conditions. The difference can be explained within different approaches. The first approach is the dog fight between players. It is observed by researches and reported by participants that there was a dog fight of gaze-cue between players to deceive other players into wrong answers. In this case, players would either immediately decide their answers to lose the track or have a deceptive dog fight. This deceptive dog fight in general is leaded by a player and tried to be unbounded by other players. Yet a further work is necessarily required to investigate the phenomenon and its parameters.

Continuous investment of attention for goal-oriented competition towards five autonomous indicators increases the mental workload. Mental work-overload decreases the performance (Kantowitz, 2000). Decreased performance may be created slower transition towards deeper reasoning stages. Since this is a repeated game with a learning period between trials, a converging ratio or intuition to equilibrium may already have been gained. In fact, guesses show that it is gained. Although, an individual which has a mental work-overload, she may not be easily satisfied positioning her indicator near others while they are following and observing her indicator. This proposition also suggests the further investigation towards all gaze-cue sharing decision process. Online simultaneous games' process of reasoning may be differentiated than offline beauty contest game (Chen et al., 2018, Müller & Schwieren, 2011). Lastly, in the shell of what was designed for all gaze-cue sharing condition, there are ball shaped indicators autonomously moving left and right on the screen. The resulting confusion of this stimulation cannot be ignored.

It is expected that anchoring occurs during the reasoning process. While no and self gaze-cue are showing no difference in terms of guess time, significantly higher guess times of all gaze-cue reflects the struggle of the underlying mechanism of the anchoring heuristics. While there is no information beyond the score page, players may anchor either the group average or the winning number, while playing the game it is expected no change for the anchor. However, continuous information flow of the all gaze-cue sharing condition may distort the anchor giving in the score page or the adjustment phase. The distortion of anchor or adjustment might delay the satisficing decision.

There is a decreasing trend of guessing times from first to last trial. Therefore, gaining an intuition between trials decreases the guessing times because it also can decrease the mental workload. There is no difference in p values reported as expected. There is an interaction effect for all factors when they are compared two by two, however, no interaction effect is found when get all together.

5.3 Discussion of Eye Movements

Three different hypotheses are declared for eye movements and three different analyses were made. There is no significant effect of gaze-cue sharing on saccade frequency, saccade amplitude and fixation duration. $H_{3,}$ $H_{4,}$ and H_5 are falsified. There are several possible explanations for? the results. The first explanation refers to mechanism of the setting. Since the eyes are the cursors of the decision, the emphasis on eye movements can deviate the natural movement and randomly varied eye movements can be detected. Another explanation is the self-paced setting of the task. Since classified movements of the eyes can be uncontrollably differentiated for the individual on the task. Therefore, average of the fixation duration and saccadic amplitude and frequency of saccade may not be enough for the eye-movement analysis subjects for seeing the expected effect. The indifference observed for guesses and eye-movements can be partly accounted by the reduced alertness due to repeated rounds of play. The reported accuracy is 0.5. Accuracy fluctuations during the experiment is not taken into account. They can be consequenced as unreliable eye-movement analyses.

5.4 Conclusion

In this study, Keynesian beauty contest game is examined for testing the gaze-cue sharing effect. An online and simultaneously played beauty contest game is created for digital environment is to carry out this investigation. Each players' eye-movements are sent to the server and streamed back to each players computer all together to share the gaze-cue of players. There are three different conditions defined: no, self and all gaze-cue sharing conditions. For no gaze-cue condition, players made their guesses looking at a screen which has a ruler and a p value in the middle of the screen, for self gaze-cue sharing each players gaze-cue is sent back to their own screen, for all gaze-cue sharing condition all gaze-cue indicators stream to each player's screen. It was expected that this gaze-cue sharing conditions have an effect on guesses that players made, guess times and eye movements. For eye-movements, fixation duration, saccade frequency and saccade amplitude are measured.

There are three different p values used in the study; "1/2", "2/3" and "4/3". It is expected that p values greater than "1" cause players to converge into "100" and smaller than one cause players to converge into "0". Since there is a task-based difference expected in the guess values depended on p values, all p values are analyzed adjusted as their distance from equilibrium and no significant difference is found for gaze-cue sharing conditions when p - values are taken together. Follow-up analysis shows that "1/2" and "4/3" is significantly different when they are examined together with gaze-cue sharing conditions. It is found that guess times are significantly higher for all gaze-cue conditions and there is a learning effect which makes guesses quicker from first trial to last. Lastly, it is found that no average fixation duration, saccade amplitude and saccade frequency difference.

Beauty contest is already a well investigated economical game. The importance of this study is its experimental setting which presents an online simultaneous experience for the participant using GET platform. The interaction of the participants using gaze-cue sharing is observed in the way which is not done before in the literature best of our knowledge. The results of the study give insights about decision making and heuristics, theory of mind and social cognition. Reached results of the study is not enough itself to enlighten the phenomena revealed but provided future work may achieve this.

5.5 Limitations of the Work and Future Directions

The present study can be accepted as a gate through different approaches to game theory. Hence, it is clear that there are some limitations can be noted and future directions to follow. The general limitations of the study are the sample size, the devices used, and experiment conduction features. There must be also follow-up studies focused on the reasoning process, more detailed eye-movement analysis and modeling the agents' behavior for all gaze-cue sharing conditions. There can be also multi-modal analysis with Functional - Near Infrared Spectroscopy (F-NIRS) and Galvanic Skin Response (GSR).

Groups consisting of five participants took part in six different experiments, amassing a total of 30 participants. Counter balance of the three conditions of gaze-cue sharing is only cycled once. The effect of priority and recency can be normalized with more cycles and thus the behavioral differences between conditions can be reduced. Since the experiment is based on social interactions, the real-life relationship between participants can have an impact on their performance. Participants were taken together to the experiment room, ignoring their relationship and familiarity prior to the experiment. Participants reported that they tried to recognize their acquaintances' indicator in the all gaze-cue sharing conditions and in the score page. They also reported that their strategy was biased to familiarity attributed indicators.

Eye-tribe is used in the experiment with a configuration of 60 Hz. data collection per participant. It is practical using the open source code of the EyeTribe to be integrated custom made scripts such as the present study does. Eye movement classification can be made with 60 Hz.. However, alternative measures to detect alertness could be used with devices. Also default scripts and statistics by the device itself.

Physiological state of the human body can be affected by the other physiological elements. The experiment conducted without controlling the the exact hour, sleep deprivation and psychological neuroactives and psychoactives used, emotional states and etc.. Also, genders are not matched for reducing the confounding effect on social group dynamics.

In the literature of economics, studies subjecting the beauty contest games focus on the reasoning process in general enhancing the model of the bounded rationality and depth of reasoning. This current study is particularly focused on the evolutionary effect of the gaze-cue sharing on the reasoning. A comparison between the models of no or self gaze-cue sharing and all gaze-cue sharing can be examined as a follow-up investigation. The computational models of the isolated version can be compared with the mathematical models of level-k and cognitive hierarchy. The simultaneous version can be modelled to imitate the discussed phenomena of satisficing decision, leading effect and herd behavior.

The eye-movement analysis made, is done to reveal the basic difference between gaze-cue sharing conditions. As the GET platform supports, eye-movement analysis can be made including smooth pursuits as well. Besides classifications, metrics such as convergence of players' gaze can be provided. Also, as an increasing branch of study, multi-modal analyses using more than one physiology monitoring can be used. It is expected that physiological aspects of anchoring can be measured using GSR. Furthermore, theory of mind and social cognition phenomena can be studied with brain to brain coherence using multiple F-NIRS. These further research of modeling and multi-modal analyses can reveal the mechanism underlying the effects found in this study.

REFERENCES

Adams, D. (1980). *The hitchhiker's guide to the galaxy*. New York, Harmony Books

Apperly, I. (2010). *Mindreaders: The cognitive basis of "theory of mind"*. Psychology Press.

Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology*, *44*(1), 1465-1467.

Bergman, O., Ellingsen T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters 107*(1), 66-68.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity, ... : Newspaper and lab beauty-contest experiments. *The American Economic Review*, *92*(5), 1687–1701.

Bühren, C., Frank, B., & Rosemarie, N. (2012). *A historical note on the beauty contest.*

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004, August). A cognitive hierarchy model of games. *The Quarterly Journal of Economics, 119*(3), 861–898.

Chapman, G. B., & Johnson E. J., (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*(1), 223-242.

Chapman, G. B., & Johnson E. J., (1999). Anchoring, activation and the construction of values. *Organizational Behavior and Human Decision Process Processes, 79*(1), 1–39.

Chen, C.-T., Huang, C.-Y., & yi Wang, J. T. (2018, September). A window of cognition: Eyetracking the reasoning process in spatial beauty contest games. *Games and Economic Behavior, 111*, 143–158.

Costa-Gomes, M., Crawford, V., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*(5), 1193-1235.

Deniz, O. (2016). Group eye tracking [Computer software manual]. METU.

Deniz, O., Fal, M., & Acatürk., C. (2015). Get-social: Group eye tracking environment for social gaze analysis. In Vienna: *The journal of eye movement.* (An optional note)

Duffy, J., & Nagel, R. (1997). On the robustness of behaviour in experimental 'beauty contest' games. *The Economic Journal, 107*(445), 1684–1700.

Epley, N., & Gillovich T. (2001). Putting adjustment back into the anchoring and adjustment heuristic:differential processing of self- generated or experimenter-provided ancors. *Psychological Science,* 12(1), 391 – 396.

Epley, N., & Gillovich T. (2005). When effortful thinking influences judgmental anchoring: differential effects of forewarning and incentives, on self-generated or externally provided anchors. *Journal of Behavioral Decision Making, 18(1),* 199-212

Forgas, J. P. (1983, June). What is social about social cognition? *British Journal of Social Psychology, 22*(2), 129–144.

Frederick, S., Kahneman, D., & Mochon, D. (2010). Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology, 20*(1), 17-19.

Frith, C. D., & Frith, U. (2012, January). Mechanisms of social cognition. *Annual Review of Psychology, 63*(1), 287–313.

Frith, C. D., & Singer, T. (2008, October). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1511), 3875–3886.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*(1), 35-42.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75–90.

Goodie, A. S., Doshi, P., & Young, D. L. (2010, October). Levels of theory of-mind reasoning in competitive games. *Journal of Behavioral Decision Making, 25*(1), 95–108.

Hedden,T., & Zhang, J. (2002,August). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition, 85*(1), 1–36.

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *The American Economic Review, 88*(4), 947–969.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kantowitz, B. H. (2000, July). Attention and mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 44*(21), 3–456.

Keynes, J. M. (1937). *The general theory of employment, interest and money*. Macmillan.

Kocher, M., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty contest games. *Economic Journal, 115*(500), 200-223.

Kocher, M., & Sutter, M. (2006, November). Time is money —Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior Organization, 61*(3), 375–392.

McKinney, W. (2010). Data structures for statistical computing in python. In S.van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (p. 51 - 56).

Meijering, B., van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2013, June). Correction: What eye movements can tell about theory of mind in a strategic game. *PLoS ONE, 8*(6).

Moulin, H. (1986*). Game theory for the social sciences*. NYU Press.

Müller, J., & Schwieren, C. (2011). More than meets the eye: an eyetracking experiment on the beauty contest game. *Mimeo*.

Nagel, R. (1993). Experimental results on interactive competitive guessing. *American Economic Review, 85*(6), 1313-1326.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review, 85*(5), 1313-26.

Nash, J. F. (1950, January). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*(1), 48–49.

Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press, Princeton.

Oechssler, J., Roider, S., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization. 72*(1), 142 - 157.

Özkul, C., M. (2018). Group eye tracking platform manual [Computer software manual]. METU.

Perner, J., Leekam, S. R., & Wimmer, H. (1987, June). Three-year-olds difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125–137.

Perner, J., & Wimmer, H. (1985). "john thinks that mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437–471.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on eye tracking research & applications*-ETRA00. ACMPress.

Simon, H. A. (1996). *The sciences of the artificial (3rd ed.).* Cambridge, MA, USA: MIT Press.

Slonim, R. L. (2005). Competing against experienced and inexperienced players. *Experimental Economics, 8*(1), 55–75.

Strack, F., & Mussweiler, T (1997). Explaining the enigmatic anchoring effect: mechanisms of selective accesibility. *Journal of Personality and Social Psychology, 73*(1), 437-446

Tajaddini, M. (2018). *Recurrence quantification analysis of group eye tracking* (Unpublished master's thesis). Middle Easy Technical University.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristic and biases*, Science*, *185*(1), 1124 - 1131.

Weber, R. A. (2003). Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior, 44*(1), 134 144.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Jarvis W.B.G. (2001). Implications of attitude change theories for numerical anchoring; anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology, 37*(1), 62 - 69.

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: implications of attitude theories for consumer judgment and decision making. *Journal of consumer psychology, 20*(1), 5 – 16.

Wilson T. D., Houston C. E., Etling K. M., & Brekke N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*(1), 387-402.

APPENDICES

APPENDIX A

Consent Form

KATILIMCI ONAM FORMU

Sayın katılımcı,

Çalışmaya katılmayı kabul ettiğiniz için teşekkür ederiz. Bu çalışma Doç. Dr. Cengiz Acartürk'ün danışmanlığında TÜBİTAK 1001 Bilimsel ve Teknolojik Araştırma Projelerini Destekleme Programı kapsamında Hazal Arpacı, Emre Erçin, Şura Genç ve Tuğçe Gürbüz tarafından yürütülmektedir. Çalışma ile ilgili aşağıdaki bilgileri lütfen dikkatlice okuyunuz ve sorularınız olursa lütfen araştırmacılara sorunuz.

Bu çalışmanın amacı tahmin ve karar verme sürecinin sosyal bağlamda göz izleme araçları ile kayıt altına alınması ve incelemesidir. Çalışma yaklaşık olarak 30 dakika sürecektir. Çalışmada rahatsızlık verici herhangi bir materyal bulunmamaktadır ancak istediğiniz zaman çalışmayı bırakabilirsiniz. Çalışma sırasında alıcı sistemler göz bebeği büyüklüğü gibi ölçümler yapmakta ve kaydetmektedir. Kaydedilen veriler gizli tutulacak ve sadece bilimsel araştırma amacıyla kullanılacaktır. Çalışmaya 5 kişilik gruplar halinde toplam 30 katılımcı dahil edilecektir. Çalışmaya katılım gönüllülük esasına dayanmaktadır ve çalışma sonunda her katılımcı ortalama 20 TL katılım bedeli alacaktır.

Çalışma sonrasında sormak istediğiniz sorular olursa ya da çalışma ile ilgili bilgi almak isterseniz e.ercin.93@gmail.com iletişim adresinden araştırmacılara ulaşabilirsiniz.

"Yukarıda yer alan çalışma ile ilgili katılımcılara verilmesi gereken bilgileri içeren metni okudum ve anladım. Çalışma ile ilgili sormak istediğim soruları araştırmacılara sordum ve gerekli cevapları aldım. Çalışmaya kendi isteğimle katılmaktayım ve çalışmadan istediğim zaman ayrılabileceğimi biliyorum. Çalışma kapsamında elde edilen şahsıma ait verilerin bilimsel araştırma amacıyla kullanılmasını, gizlilik kurallarına uyulmak kaydıyla sunulmasını ve yayınlanmasını kabul ettiğimi beyan ederim."

Katılımcının adı/soyadı: Araştırmacının/ların adı/soyadı:

İmza/Tarih: İmza/Tarih:

APPENDIX B

Demographical Information Form

DEMOGRAFİK BİLGİ FORMU

Katılımcı Numarası: ____________

Demografik Bilgiler

Doğum yılı: _____________________

Cinsiyet: _____________________

Eğitim durumu: ☐ İlkokul ☐ Ortaokul ☐ Lise ☐ Lisans ☐ Yüksek lisans ☐ Doktora

Okul: ________________________

Bölüm: ________________________

Sınıf: ________________________

Genel Bilgiler

Gözlük ya da lens kullanıyor musunuz?

☐ Evet

☐ Hayır

Herhangi bir psikolojik/nörolojik rahatsızlığınız var mı?

☐ Evet (Belirtiniz:____)

☐ Hayır

Herhangi bir ilaç kullanıyor musunuz?

☐ Evet (Belirtiniz:____)

☐ Hayır

APPENDIX C

Instructions

Sizinle birlikte çalışmaya giren diğer katılımcılarla bir tahmin oyunu oynayacaksınız. Amacınız grup tahmin ortalamasının verilen katsayı ile çarpımına en yakın tahmini yapmak. Örneğin; grup tahmin ortalamasının X olduğu ve verilen katsayının P olduğu durumda, X*P'ye en yakın tahmini yapan yani en düşük hata puanını alan oyuncu turu kazacanak. Oyun sonunda alacağınız para ödülü kazandığınız tur sayısı ile değişecek.

Tahminlerinizi yapmak için aşağıdaki ekranı göreceksiniz:

Cetvelin üzerindeki kesirli sayı hedef tahmin için kullanılan katsayıyı belirtmektedir. Bulunduğunuz koşula göre katsayıyla birlikte yalnızca cetveli, cetvelin üzerinde yalnızca kendi göz izinizi veya tüm katılımcıların göz izini görebilirsiniz. Tahmininizi yapmak için cetvelde tahmin edeceğiniz sayıya bakmalı ve "BOŞLUK" ("SPACE") tuşuna basmalısınız. "BOŞLUK" tuşuna bastığınız anda göz iziniz cetvelde üzerinde donacak, tahmininiz göz izinin üzerinde sayılı olarak belirtilecek ve karar evresine geçeceksiniz. Karar evresine geçtikten sonra diğer katılımcılarla ilişkiniz kesilecek ve bu soru içinde tahmin evresine dönemeyeceksiniz. Eğer göz iziniz ve dolayısıyla seçiminiz tam olarak yapmak istediğiniz tahminin üzerinde donmadıysa, karar evresinde tahmnininizi beş birim kadar artırabilir veya azaltabilirsiniz.

Karar evresinde aşağıdaki ekranı göreceksiniz:

"SOL" tuşuna bastığınızda tahmninizi en fazla beş birime kadar azaltabilirsiniz. "SAĞ" tuşa bastığınızda ise tahmininizi yine en fazla beş birime kadar arttırabilirsiniz. "AŞAĞI" tuşuna bastığınızda ise son kararınızı vermiş olursunuz ve diğer katılımcıların son kararlarını vermesini beklersiniz. Tahmin ve karar aşamalarında herhangi bir süre sınırı bulunmamaktadır.

APPENDIX D

Eye Movement Classifier based on I-VT Filter

```
data = pd.read_csv("individualgazelog/" + file, low_memory = False)
lastx = 0
lasty = 0
laststs = 0
h = 30
d = 56
r = 768
movement= []
for row in data.iterrows():
if ',' in row[1]['coordinates']:
x = float(row[1]['coordinates'].split(',', 1)[0])
y = float(row[1]['coordinates'].split(',', 1)[1])
sts = int(row[1]['sts'])
size_in_px = math.sqrt((x-lastx)**2+(y-lasty)**2)
deg_per_px = degrees(atan2(.5*h, d)) / (.5*r)
size_in_deg = size_in_px * deg_per_px
if sts-laststs == 0 or x == lastx or y == 0:
velocity = 0
else:
velocity = size_in_deg/((sts-laststs)/1000)
lastx = x
lasty = y
laststs = sts
if velocity < 100:
movement.append('fixation')
elif velocity > 300:
movement.append('saccade')
else:
movement.append('mixed')
else:
movement.append('')
data['eyemovement'] = movement
```

APPENDIX E

List of the Tables

Table E4.1

Within subject effects of distance from equilibrium of guesses for gaze-cue conditions, p-values and trials

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Gaze-Cue	None	7.013e +10[a]	2.000[a]	3.506e +10[a]	45.093[a]	1.535e -12[a]	0.609
	Greenhouse-Geisser	7.013e +10[a]	1.058[a]	6.628e +10[a]	45.093[a]	1.140e -7[a]	0.609
Residual	None	4.510e +10	58.000	7.776e +8			
	Greenhouse-Geisser	4.510e +10	30.684	1.470e +9			
P-Value	None	1.128e +9	2.000	5.642e +8	3.033	0.056	0.095
	Greenhouse-Geisser	1.128e +9	1.955	5.772e +8	3.033	0.057	0.095
Residual	None	1.079e +10	58.000	1.860e +8			
	Greenhouse-Geisser	1.079e +10	56.699	1.903e +8			
Trials	None	1.038e +10[a]	3.000[a]	3.459e +9[a]	10.688[a]	4.742e -6[a]	0.269
	Greenhouse-Geisser	1.038e +10[a]	1.563[a]	6.638e +9[a]	10.688[a]	4.513e -4[a]	0.269
Residual	None	2.816e +10	87.000	3.237e +8			
	Greenhouse-Geisser	2.816e +10	45.335	6.211e +8			
Gaze-Cue* P-Value	None	3.296e +9[a]	4.000[a]	8.240e +8[a]	4.338[a]	0.003[a]	0.130
	Greenhouse-Geisser	3.296e +9[a]	2.111[a]	1.561e +9[a]	4.338[a]	0.016[a]	0.130
Residual	None	2.204e +10	116.000	1.900e +8			
	Greenhouse-Geisser	2.204e +10	61.227	3.599e +8			
Gaze-Cue* Trials	None	6.036e +9[a]	6.000[a]	1.006e +9[a]	4.170[a]	6.081e -4[a]	0.126
	Greenhouse-Geisser	6.036e +9[a]	1.566[a]	3.855e +9[a]	4.170[a]	0.030[a]	0.126
Residual	None	4.198e +10	174.000	2.413e +8			
	Greenhouse-Geisser	4.198e +10	45.407	9.245e +8			
P-Value* Trials	None	3.356e +9[a]	6.000[a]	5.594e +8[a]	3.255[a]	0.005[a]	0.101
	Greenhouse-Geisser	3.356e +9[a]	2.507[a]	1.339e +9[a]	3.255[a]	0.034[a]	0.101
Residual	None	2.990e +10	174.000	1.718e +8			
	Greenhouse-Geisser	2.990e +10	72.710	4.112e +8			
Gaze-Cue* P-Value* Trials	None	5.082e +9[a]	12.000[a]	4.235e +8[a]	2.666[a]	0.002[a]	0.084
	Greenhouse-Geisser	5.082e +9[a]	2.333[a]	2.178e +9[a]	2.666[a]	0.068[a]	0.084
Residual	None	5.528e +10	348.000	1.589e +8			
	Greenhouse-Geisser	5.528e +10	67.662	8.171e +8			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Table E4.2

Holm's Correction Post-Hoc Test of distance from equilibrium of guesses for p-values

Post Hoc Comparisons - P-Value

		Mean Difference	SE	t	Cohen's d	p holm
1/2	2/3	-1.856	2.716	-0.683	-0.125	0.497
	4/3	-16.447	2.716	-6.056	-1.106	3.309e-7
2/3	4/3	-14.592	2.716	-5.372	-0.981	2.872e-6

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons

Table E4.3

Holm's Correction Post-Hoc Test of distance from equilibrium of guesses for trials

Post Hoc Comparisons - Trials

		Mean Difference	SE	t	Cohen's d	p holm
1	2	11.296	1.354	8.341	1.523	3.998e-12
	3	14.911	1.354	11.010	2.010	1.805e-17
	4	18.437	1.354	13.614	2.486	1.707e-22
2	3	3.615	1.354	2.669	0.487	0.018
	4	7.141	1.354	5.273	0.963	2.905e-6
3	4	3.526	1.354	2.604	0.475	0.018

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons

Table E4.4

Within subject effects of guess time for gaze-cue conditions, p-values and trials

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Gaze-Cue	None	7.013e +10[a]	2.000[a]	3.506e +10[a]	45.093[a]	1.535e -12[a]	0.609
	Greenhouse-Geisser	7.013e +10[a]	1.058[a]	6.628e +10[a]	45.093[a]	1.140e -7[a]	0.609
Residual	None	4.510e +10	58.000	7.776e +8			
	Greenhouse-Geisser	4.510e +10	30.684	1.470e +9			
P-Value	None	1.128e +9	2.000	5.642e +8	3.033	0.056	0.095
	Greenhouse-Geisser	1.128e +9	1.955	5.772e +8	3.033	0.057	0.095
Residual	None	1.079e +10	58.000	1.860e +8			
	Greenhouse-Geisser	1.079e +10	56.699	1.903e +8			
Trials	None	1.038e +10[a]	3.000[a]	3.459e +9[a]	10.688[a]	4.742e -6[a]	0.269
	Greenhouse-Geisser	1.038e +10[a]	1.563[a]	6.638e +9[a]	10.688[a]	4.513e -4[a]	0.269
Residual	None	2.816e +10	87.000	3.237e +8			
	Greenhouse-Geisser	2.816e +10	45.335	6.211e +8			
Gaze-Cue* P-Value	None	3.296e +9[a]	4.000[a]	8.240e +8[a]	4.338[a]	0.003[a]	0.130
	Greenhouse-Geisser	3.296e +9[a]	2.111[a]	1.561e +9[a]	4.338[a]	0.016[a]	0.130
Residual	None	2.204e +10	116.000	1.900e +8			
	Greenhouse-Geisser	2.204e +10	61.227	3.599e +8			
Gaze-Cue* Trials	None	6.036e +9[a]	6.000[a]	1.006e +9[a]	4.170[a]	6.081e -4[a]	0.126
	Greenhouse-Geisser	6.036e +9[a]	1.566[a]	3.855e +9[a]	4.170[a]	0.030[a]	0.126
Residual	None	4.198e +10	174.000	2.413e +8			
	Greenhouse-Geisser	4.198e +10	45.407	9.245e +8			
P-Value* Trials	None	3.356e +9[a]	6.000[a]	5.594e +8[a]	3.255[a]	0.005[a]	0.101
	Greenhouse-Geisser	3.356e +9[a]	2.507[a]	1.339e +9[a]	3.255[a]	0.034[a]	0.101
Residual	None	2.990e +10	174.000	1.718e +8			
	Greenhouse-Geisser	2.990e +10	72.710	4.112e +8			
Gaze-Cue* P-Value* Trials	None	5.082e +9[a]	12.000[a]	4.235e +8[a]	2.666[a]	0.002[a]	0.084
	Greenhouse-Geisser	5.082e +9[a]	2.333[a]	2.178e +9[a]	2.666[a]	0.068[a]	0.084
Residual	None	5.528e +10	348.000	1.589e +8			
	Greenhouse-Geisser	5.528e +10	67.662	8.171e +8			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table E4.5

Holm's Correction Post-Hoc Test of guess times for gaze-cue conditions

Post Hoc Comparisons - Gaze-Cue

		Mean Difference	SE	t	Cohen's d	p_{holm}
All	None	17066.558	2078.471	8.211	1.499	7.356e -11
	Self	17121.403	2078.471	8.237	1.504	7.356e -11
None	Self	54.844	2078.471	0.026	0.005	0.979

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons

Table E4.6

Holm's Correction Post-Hoc Test of guess times for trials

Post Hoc Comparisons - Trials

		Mean Difference	SE	t	Cohen's d	p_{holm}
1	2	3614.996	1548.358	2.335	0.426	0.066
	3	6759.585	1548.358	4.366	0.797	1.739e-4
	4	8011.830	1548.358	5.174	0.945	8.704e-6
2	3	3144.589	1548.358	2.031	0.371	0.091
	4	4396.833	1548.358	2.840	0.518	0.022
3	4	1252.244	1548.358	0.809	0.148	0.421

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons.

Table E4.7

Within subject effects of fixation duration for gaze-cue conditions, p-values and trials

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Gaze-Cue	None	877314.015	2.000	438657.007	2.200	0.120	0.071
	Greenhouse-Geisser	877314.015	1.729	507369.196	2.200	0.128	0.071
Residual	None	1.156e+7	58.000	199347.048			
	Greenhouse-Geisser	1.156e+7	50.145	230573.204			
P-Value	None	120295.221	2.000	60147.611	0.762	0.471	0.026
	Greenhouse-Geisser	120295.221	1.993	60372.760	0.762	0.471	0.026
Residual	None	4.576e+6	58.000	78896.562			
	Greenhouse-Geisser	4.576e+6	57.784	79191.895			
Trials	None	789824.834[a]	3.000[a]	263274.945[a]	3.517[a]	0.018[a]	0.108
	Greenhouse-Geisser	789824.834[a]	1.848[a]	427366.862[a]	3.517[a]	0.040[a]	0.108
Residual	None	6.512e+6	87.000	74848.691			
	Greenhouse-Geisser	6.512e+6	53.595	121499.789			
Gaze-Cue* P-Value	None	984994.752	4.000	246248.688	2.891	0.025	0.091
	Greenhouse-Geisser	984994.752	3.319	296805.830	2.891	0.034	0.091
Residual	None	9.881e+6	116.000	85179.957			
	Greenhouse-Geisser	9.881e+6	96.241	102668.193			
Gaze-Cue* Trials	None	453901.956[a]	6.000[a]	75650.326[a]	1.118[a]	0.354[a]	0.037
	Greenhouse-Geisser	453901.956[a]	4.527[a]	100270.668[a]	1.118[a]	0.353[a]	0.037
Residual	None	1.177e+7	174.000	67648.281			
	Greenhouse-Geisser	1.177e+7	131.276	89664.364			
P-Value* Trials	None	491689.490[a]	6.000[a]	81948.248[a]	1.618[a]	0.145[a]	0.053
	Greenhouse-Geisser	491689.490[a]	4.562[a]	107768.171[a]	1.618[a]	0.165[a]	0.053
Residual	None	8.811e+6	174.000	50640.547			
	Greenhouse-Geisser	8.811e+6	132.312	66596.165			
Gaze-Cue* P-Value* Trials	None	1.151e+6[a]	12.000[a]	95922.341[a]	1.448[a]	0.142[a]	0.048
	Greenhouse-Geisser	1.151e+6[a]	6.473[a]	177823.287[a]	1.448[a]	0.194[a]	0.048
Residual	None	2.306e+7	348.000	66256.083			
	Greenhouse-Geisser	2.306e+7	187.720	122827.221			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table E4.8
Holm's Correction Post-Hoc Test of fixation duration for trials

Post Hoc Comparisons - Trials

		Mean Difference	SE	t	Cohen's d	p_{holm}
1	2	26.165	23.546	1.111	0.203	0.539
	3	58.033	23.546	2.465	0.450	0.078
	4	68.661	23.546	2.916	0.532	0.027
2	3	31.868	23.546	1.353	0.247	0.538
	4	42.496	23.546	1.805	0.330	0.298
3	4	10.628	23.546	0.451	0.082	0.653

Note. Cohen's d does not correct for multiple comparisons.
Note. Bonferroni adjusted confidence intervals.

Table E4.9

Within subject effects of saccade frequency for gaze-cue conditions, p-values and trials

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Gaze-Cue	None	1.834[a]	2.000[a]	0.917[a]	0.622[a]	0.540[a]	0.021
	Greenhouse-Geisser	1.834[a]	1.622[a]	1.131[a]	0.622[a]	0.509[a]	0.021
Residual	None	85.525	58.000	1.475			
	Greenhouse-Geisser	85.525	47.041	1.818			
P-Value	None	19.838[a]	2.000[a]	9.919[a]	2.136[a]	0.127[a]	0.069
	Greenhouse-Geisser	19.838[a]	1.119[a]	17.731[a]	2.136[a]	0.152[a]	0.069
Residual	None	269.313	58.000	4.643			
	Greenhouse-Geisser	269.313	32.447	8.300			
Trials	None	2.125[a]	3.000[a]	0.708[a]	1.271[a]	0.289[a]	0.042
	Greenhouse-Geisser	2.125[a]	2.221[a]	0.957[a]	1.271[a]	0.289[a]	0.042
Residual	None	48.498	87.000	0.557			
	Greenhouse-Geisser	48.498	64.416	0.753			
Gaze-Cue* P-Value	None	2.944[a]	4.000[a]	0.736[a]	0.628[a]	0.643[a]	0.021
	Greenhouse-Geisser	2.944[a]	1.546[a]	1.904[a]	0.628[a]	0.499[a]	0.021
Residual	None	135.885	116.000	1.171			
	Greenhouse-Geisser	135.885	44.832	3.031			
Gaze-Cue* Trials	None	2.960[a]	6.000[a]	0.493[a]	0.789[a]	0.579[a]	0.026
	Greenhouse-Geisser	2.960[a]	2.910[a]	1.017[a]	0.789[a]	0.500[a]	0.026
Residual	None	108.734	174.000	0.625			
	Greenhouse-Geisser	108.734	84.402	1.288			
P-Value* Trials	None	1.704[a]	6.000[a]	0.284[a]	0.598[a]	0.732[a]	0.020
	Greenhouse-Geisser	1.704[a]	3.483[a]	0.489[a]	0.598[a]	0.642[a]	0.020
Residual	None	82.628	174.000	0.475			
	Greenhouse-Geisser	82.628	101.002	0.818			
Gaze-Cue* P-Value* Trials	None	8.985[a]	12.000[a]	0.749[a]	1.335[a]	0.197[a]	0.044
	Greenhouse-Geisser	8.985[a]	3.019[a]	2.976[a]	1.335[a]	0.268[a]	0.044
Residual	None	195.215	348.000	0.561			
	Greenhouse-Geisser	195.215	87.562	2.229			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table E4.10

Within subject effects of saccade amplitude for gaze-cue conditions, p-values and trials

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Gaze-Cue	None	136.598	2.000	68.299	1.815	0.172	0.059
	Greenhouse-Geisser	136.598	1.843	74.105	1.815	0.175	0.059
Residual	None	2181.985	58.000	37.620			
	Greenhouse-Geisser	2181.985	53.456	40.819			
P-Value	None	173.236[a]	2.000[a]	86.618[a]	4.024[a]	0.023[a]	0.122
	Greenhouse-Geisser	173.236[a]	1.631[a]	106.221[a]	4.024[a]	0.032[a]	0.122
Residual	None	1248.425	58.000	21.525			
	Greenhouse-Geisser	1248.425	47.296	26.396			
Trials	None	152.921	3.000	50.974	3.224	0.026	0.100
	Greenhouse-Geisser	152.921	2.706	56.512	3.224	0.031	0.100
Residual	None	1375.510	87.000	15.810			
	Greenhouse-Geisser	1375.510	78.473	17.528			
Gaze-Cue* P-Value	None	115.555	4.000	28.889	1.681	0.159	0.055
	Greenhouse-Geisser	115.555	3.465	33.352	1.681	0.168	0.055
Residual	None	1993.418	116.000	17.185			
	Greenhouse-Geisser	1993.418	100.477	19.839			
Gaze-Cue* Trials	None	135.960	6.000	22.660	1.268	0.274	0.042
	Greenhouse-Geisser	135.960	4.300	31.617	1.268	0.285	0.042
Residual	None	3109.290	174.000	17.869			
	Greenhouse-Geisser	3109.290	124.707	24.933			
P-Value* Trials	None	153.389	6.000	25.565	1.527	0.172	0.050
	Greenhouse-Geisser	153.389	5.001	30.671	1.527	0.185	0.050
Residual	None	2913.864	174.000	16.746			
	Greenhouse-Geisser	2913.864	145.034	20.091			
Gaze-Cue* P-Value* Trials	None	223.499	12.000	18.625	1.106	0.354	0.037
	Greenhouse-Geisser	223.499	7.486	29.856	1.106	0.360	0.037
Residual	None	5858.981	348.000	16.836			
	Greenhouse-Geisser	5858.981	217.091	26.989			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table E4.11

Holm's Correction Post-Hoc Test of saccade amplitude for p value conditions

Post Hoc Comparisons - P-Value

		Mean Difference	SE	t	Cohen's d	p_{holm}
1/2	2/3	-0.915	0.346	-2.645	-0.483	0.031
	4/3	-0.150	0.346	-0.434	-0.079	0.666
2/3	4/3	0.765	0.346	2.211	0.404	0.062

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons

Table E4.12

Holm's Correction Post-Hoc Test of saccade amplitude for trials

Post Hoc Comparisons - Trials

		Mean Difference	SE	t	Cohen's d	p_{holm}
1	2	0.934	0.342	2.730	0.498	0.046
	3	0.385	0.342	1.124	0.205	0.536
	4	0.849	0.342	2.480	0.453	0.075
2	3	-0.549	0.342	-1.606	-0.293	0.448
	4	-0.085	0.342	-0.250	-0.046	0.803
3	4	0.464	0.342	1.356	0.248	0.536

Note. Bonferroni adjusted confidence intervals.
Note. Cohen's d does not correct for multiple comparisons