THE IMPACT OF MODALITY AND FEEDBACK ON REASONING ABOUT BASE RATE NEGLECT PROBLEMS IN BEHAVIORAL AND EYE TRACKING STUDIES: A COGNITIVE SCIENCE PERSPECTIVE

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

BY

BURCU VERİM

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

SEPTEMBER 2014

THE IMPACT OF MODALITY AND FEEDBACK ON REASONING ABOUT BASE RATE NEGLECT PROBLEMS IN BEHAVIORAL AND EYE TRACKING STUDIES: A COGNITIVE SCIENCE PERSPECTIVE

submitted by **BURCU VERİM** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

Prof. Dr. Nazife Baykal _____

Director, **Informatics Institute**

Prof. Dr. Cem Bozşahin _____

Head of Department, **Cognitive Science**

Assoc. Prof. Dr. Annette Hohenberger _____

Supervisor, **Cognitive Science, METU**

Examining Committee Members:

Prof. Dr. Cem Bozşahin _____

Cognitive Science, METU

Assoc. Prof. Dr. Annette Hohenberger _____

Cognitive Science, METU

Assist. Prof. Dr. Cengiz Acartürk _____

Cognitive Science, METU

Assist. Prof. Dr. Murat Perit Çakır _____

Cognitive Science, METU

Prof. Dr. Dilek Önkal _____

Business Administration, Bilkent University

Date: 03.09.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I have fully cited and referenced all materials and results that are not original to this work.

Name, Last name : **Burcu Verim**

Signature : **_____**

ABSTRACT

THE IMPACT OF MODALITY AND FEEDBACK ON REASONING ABOUT BASE RATE NEGLECT PROBLEMS IN BEHAVIORAL AND EYE TRACKING STUDIES: A COGNITIVE SCIENCE PERSPECTIVE

Verim, Burcu

Master of Science, Department of Cognitive Science

Supervisor: Assoc. Prof. Dr. Annette Hohenberger

September 2014, 73 pages

The base rate fallacy is a type of reasoning error which is rooted in judgements of humans about the likelihood of some state or some event on prior beliefs and intuitions about the representativeness of the problem while neglecting base rate probabilities of this state or event. So far, base rate neglect problems have been presented in the verbal modality, in the form of story scripts. The purpose of this study is to investigate to what extent the different modalities (verbal vs. graphical) have an effect on reasoning about these problems via providing random sampling and feedback. Eye-tracking data, reaction times and accuracy rates of judgments will be measured. It is hypothesized that providing the participants feedback for questions about frequency distributions along with the direct experience of random sampling will cause a decrease in the base-rate neglect. The results indicated a significant difference between the three experimental groups. This study has implications for education in terms of multi-modal teaching, learning, and reasoning.

Keywords: judgment under uncertainty, base-rate neglect, feedback, eye-tracking

ÖZ

BİLİŞSEL BİLİMLER PERSPEKTİFİNDEN DAVRANIŞSAL ÇALIŞMALARDA VE GÖZ İZLEME ÇALIŞMALARINDA MODALİTENİN VE GERİ BİLDİRİMİN TEMEL ORAN YANILGISI PROBLEMLERİ ÜZERİNE DÜŞÜNMEDEKİ ETKİSİ

Verim, Burcu

Master, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Doç. Dr. Annette Hohenberger

Eylül 2014, 73 sayfa

Temel oran yanılgısı problemleri, temelinde insanların problemlerin temsili hakkında bir durumun ya da bir olayın olma olasılığını değerlendirirken o durum ya da olay ile ilgili temel oran olasılıklarını görmezden gelerek önceden sahip oldukları inanç ve sezgilerine dayanma yatan, bir çeşit düşünme hatasıdır. Şu ana kadar, temel oran yanılgısı problemleri sözel modalite ile senaryolaştırılmış hikayeler şeklinde sunulmuştur. Bu çalışmanın amacı, katılımcılara rastgele örnekleme fikri ve geri bildirim vererek farklı modalitelerin (sözel ve grafiksel) bu problemler üzerine düşünmeyi ne derece etkilediğini araştırmaktır. Göz-izleme verileri, reaksiyon zamanları ve yargıların doğruluk oranları ölçülecektir. Katılımcılara bir örneklemden rastgele bir kişi seçme deneyimini sağlamakla birlikte frekans dağılımlarıyla ilgili sorular için geri bildirim vermenin temel oran yanılgısında düşüşe sebep olacağı varsayılmaktadır. Sonuçlar üç deneysel grup arasında anlamlı bir farklılık olduğunu göstermektedir. Bu çalışma eğitimde çok modlu öğretim, öğrenme ve akıl yürütme alanları ile bağlantılıdır.

Anahtar Kelimeler: belirsizlik altında karar verme, temel oran yanılgısı, geri bildirim, göz izleme

To my father,

To my mother,

And to my best friend

ACKNOWLEDGMENTS

I would like to express my greatest appreciation to Assoc. Prof. Dr. Annette Hohenberger for the continuous support throughout my Master's study, for her enthusiasm, motivation and patience. I also would like to thank Assist. Prof. Dr. Cengiz Acartürk, Assist. Prof. Dr. Murat Perit Çakır and the other members of the thesis committee for their advices and constructive feedbacks.

I would like to acknowledge the financial support of TÜBİTAK (The Scientific and Technological Research Council of Turkey) during my undergraduate (Undergraduate Basic Science Scholarship (2006-2011)) and graduate studies (National Scholarship Program for MSc Students (2011-2013)).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Base-rate neglect was defined by Kahneman and Tversky (1973) as underweighting or neglecting the base-rates and relying on the representativeness heuristic when making judgments under uncertainty. Base-rate in this context means how likely it is that the behaviour or trait under consideration occurs in the population in the first place. Depending on existing beliefs to make judgments under uncertainty is relatively effortless, fast and practical in many situations. On the other hand, it causes some computational bias (Stanovich & West, 2003). In the past, many studies argued over extent to which human beings are rational (Cooper, 1989; Evans & Over, 1996; Stanovich & West, 2003). The main reason of this discussion is the finding of numerous studies demonstrating humans' inability to reason as properly as probability theories suggests. For example, it has been claimed many times that the human mind does not work in line with Bayes' Theorem. According to the Bayes' rule, existing knowledge must be updated in consideration of recently acquired evidence (Welsh and Navarro, 2012). In addition, it is evident that inhibiting the improper information is a really hard work (De Neys & Franssens, 2009). As it was the case in the current study and in many other base-rate neglect studies, Bayesian priors may not be based on empirical data; they may also based on other factors such as innate factors. That is, not all evidence is equal: some evidence is empirical (i.e., changeable) and some is innate (i.e., less or not changeable at all). This may have consequences for the updating of the posteriors, for example, the weigths that are given to the evidence may change. In this study, the evidence the participants encountered in the base-rate problems, the distributions related to attitudes, the behaviors or other characteristic features of groups of people were empirical.

Previous studies showed that people tended to ignore or significantly underweight base-rates when they did not perceive them as causal (Ajzen, 1977; Bar-Hillel, 1980; Tversky and Kahneman, 1980; Locksley and Stangor, 1984), and strongly associated with the diagnostic information/description (Kahneman & Tversky, 2000). In addition, time-pressure in decision-making tasks was found to be disadvantageous (Baddley, 1972; Hockey, 1993; Janis & Mann, 1977) because it evokes heuristic processes (Goodie & Crooks, 2004).

On the other hand, over past the forty years, researchers investigated varying methods differing in their effectiveness in order to answer the question "How can humans overcome the biases or cognitive illusions in reasoning?" (Fischhoff, 1982; Parmley, 2006). These methods comprised teaching participants how to solve problems related to

the topics probability and statistics (Kosonen & Winne, 1995), giving them instructions written in a booklet on how to do the computations with Bayesian reasoning (Nisbett *et al.,* 1982) giving participants information on the subjects of cognitive heuristics and biases (Nisbett *et al*., 1986; Bjork *et al*., 2006), presenting subjects with varying base-rates (Fischhoff, Slovic, and Lichtenstein. 1979; Birnbaum & Mellers. 1983), and conducting experiments with items in repetitive format (Kahneman & Tversky, 1973). Moreover, situations were studied where the sample in the studies that the base-rate is given from is representative (Wells & Harvey, 1977); likewise, situations where diagnostic information is not reliable (Schwarz *et al.,* 1991) and not indicative enough (Ginossar & Trope, 1980). Lastly, the use of different formats – presenting the problem using frequency instead of probability information (Gigerenzer & Hoffrage, 1995) – and participants' direct experience of random sampling (Gigerenzer, Hell, and Blank, 1988) cause a significant facilitation in reasoning about base-rate problems.

Another line of research, namely eye-tracking studies along with other behavioural experiments provided insight into the on-line processes underlying human reasoning. For instance, Ball *et al*. (2006) investigated the eye-movements of participants when they were reasoning about deductive syllogisms. These types of problems require analytic thinking as is the case with base-rate problems. There were two conditions in these problems: in one case the believability of the conclusion was congruent with the logical validity of the problem; in the second case the believability of the conclusion was incongruent with the logical validity of the problem (De Neys & Glumicic, 2008). Their results revealed longer response times for the incongruent condition. Further findings came from the experiments of De Neys and Glumicic (2008). They observed when people were trying to solve conflict and neutral base-rate problems, they needed more time to inspect either because there was a conflict between the diagnostic information and prior probabilities or there was no clue to the answer in the descriptions.

The overarching research question of the thesis was to find out how to facilate base-rate neglect problems such that subjects would understand and answer them better. In the light of preceding studies, the main reason of conducting the current thesis was to examine the impact of different modes of presentation (verbal, graphical), random sampling and feedback on base-rate neglect problems. Accuracy rates of judgments, response times and eye-tracking data (which comprised total fixation duration, total visit duration, average visit duration and visit count) were analyzed.

Before the main study, two preparatory studies were conducted, (1) a rating study and (2) a pilot study. Firstly, in order to create suitable scenarios for base-rate problems, a group of participants rated the likelihood of categories to which the protagonists (the person whose individuating information was given) belongs to. Then, 18 base-rate problems, 6 from each condition (no-conflict, conflict, neutral) were chosen and presented in each experimental condition. Secondly, a pilot study was conducted with the purpose to examine the effects of different modalities (verbal vs. graphical) on reasoning about base-rate problems. In particular, it aimed to find out if graphical presentation of the crucial probability information, prior to or simultaneously with the verbal problem statement, improved proper reasoning about base-rate problems in comparison to the classical verbal-only presentation mode.

Hypotheses for the pilot study were as follows:

H1: The graphical presentation of the crucial probability information, prior to or simultaneously with the verbal problem statement, will improve proper reasoning about base-rate problems in comparison to the classical verbal-only presentation mode.

H2: The graphical presentation of the crucial probability information, prior to or simultaneously with the verbal problem statement, will improve proper reasoning about base-rate problems in comparison to the irrelevant picture condition.

The pilot studies led to the main study by providing clues about changing the structure of the graphical presentation of the crucial probability information. The purpose of conducting the main study was to find out whether understanding the nature of random sampling, presenting the crucial probability information in a frequency format rather than as a probability statement along with feedback for the judgments will help solving base-rate problems. All participants made judgments about eighteen base-rate problems (six no-conflict, six conflict, six neutral) in three experimental groups (No feedback; Feedback; Double-Feedback). First, they all chose an individual from a sample space composed of ten neutral faces. Second, they were presented with the frequency distribution of the population indicated by coloured circles and they answered the question to which colour group the previously chosen individual was more likely to belong. Finally, they were presented with the base-rate problems and the distribution of the population together and they made judgments about the answer category to which the protagonists in the descriptions were more likely to belong.

Hypotheses of the main study that are related to accuracy scores and response times:

H1: Giving the participants the choice of choosing an individual randomly from a sample and providing them with feedback for questions about frequency distributions will cause a decrease in the base-rate neglect.

H2: Reaction times will be higher for the base-rate problems in the conflict condition than no-conflict and neutral conditions.

H3: For scenario questions related to the base-rate, accuracy rates of judgments will be the highest in the double feedback group and somewhat lower in the color feedback group. The no feedback group will be the least successful one in terms of the accuracy of answers to the scenario questions.

H4: For scenario questions, the accuracy rates of judgments for the base-rate problems in the no-conflict condition will be higher than in the neutral condition and the conflict condition will be the least successful condition for subjects.

H5: For color questions related to the crucial frequency information, the accuracy rates of judgments for the base-rate problems in the color feedback group will be higher than in the double feedback group and the no feedback group will be the least successful one.

H6: In the double feedback group, the accuracy rates of judgments will be higher in the test part than the training part.

Hypotheses of the main study that are related to eye-tracking measures:

H7: The number of switches between the AOIs (graphic, text) will be lower for the base-rate problems in the no-conflict condition than neutral condition and conflict condition will have the most number of switches.

H8: In the graphic area, total fixation duration, total visit duration and visit count will be longer for the base-rate problems in the conflict condition than in the the text area.

CHAPTER 2

LITERATURE REVIEW

2.1 Evolutionary Perspectives on Human Reasoning and Judgment Under Uncertainty

In his work on evolution of reasoning Cummins (2004) refers to human cognition as an adaptive mechanism that processes data gathered from both social and physical environments. Moreover he says that we do not encounter this information in terms of single event probabilities and odds/ratios. Instead we are exposed to it in different forms such as frequencies of entities and events. It does not go far back that probabilities and percentages have come into view as in the forms of quantitative representations in the work on evolution (Hacking, 1975). Percentages were started to be studied under the subject of uncertainty in the nineteenth century (Gigerenzer, Swijtink, Porter, Daston, Beatty, & Krüger, 1989). It seems like it took the literature a very long time to enhance the percentage and probability concepts culturally. Thus, this might be the reason why it is hard for people to use and enhance them when they reason. Especially when studying judgment under uncertainty, as a comment to Koehler (1996)'s article in which he discussed if the Bayes' Theorem was the only proper source to compare the human judgments to, Ginzburg *et al*. (1996) claimed that the reason why individuals are not good at calculating probability may be because it is not favourable in terms of evolution. In their point of view, as humans we're evolved in an environment that is autocorrelated in time and space and because judging probabilities relies on the independence hypothesis it may reveal maladaptive behaviour. They argue that the errors that our reasoning produces may not be considered as errors in terms of fuzzy arithmetic which is an area interested in a nonprobabilistic calculus of reasoning under uncertainty which consists of irrelevance of independence assumptions, sensitivity to sample sizes, neglect of prior probabilities and conservatism about uncertainty. This argument may be plausible if we consider some of the following characteristics of probability theory. First, it is not intuitive. Second as we mentioned above, there is dependence in nature, it is not like gambling devices whose structure rely on statistical independence. Third, Ginzburg *et al*. (1996) claimed that spatial patterns include positive autocorrelations. In the early ages, when humans foraged, they had to represent the next place to look for food which probably was somewhere near to the previous location. This situation is similar in animals as well (Stephan & Krebs 1986, pp. 81-90). Another support to these ideas comes from Melone & McGuire (1992). They argue that because ambiguous and uncertain cues occur frequently in our environment, non-Bayesian strategies such as cue-averaging may be favoured.

From this perspective, Koehler does not seem to be on the right way by comparing our way of making statistical judgments to Bayesian probability theory although there is no clear evidence that one theory is more plausible than the other.

2.2 Heuristics and Biases Approach

As humans we encounter many problems to solve and we have a lot of decisions to make in our daily lives. It does not seem optimal for us to take all the elements into consideration that form the problem or the situation in order to make decisions. Both problem-solving and decision-making processes have to be done quickly and effortlessly. Kahneman and Tversky (Kahneman & Tversky, 1972; Tversky & Kahneman, 1973, 1974) suggested that judgments that we make intuitively under uncertainty are generally controlled by judgmental "heuristics" rather than by the formal laws of probability. A heuristic serves the purpose of coming up with a solution in a short while without exerting conscious effort. It helps us to function in a way that we do not need to stop and think about what our next step is going to be. Heuristics are helpful to us in many circumstances for sure but meanwhile they make us prone to errors in judgment and decision-making, they lead to biases. In addition counting on past experiences and heuristics prevent us from taking into account other options or thinking of novel ideas. Kahneman and Tversky were not the first ones to challenge the models of rationality but their work thoroughly influenced many research areas such as psychology, economics, political science, medical decision making, and legal studies (Griffin *et al.* 2001). Kahneman and Tversky's initial studies indicated that intuitions made people only rely on the specific descriptions with total neglect of crucial probability information. However, in their following works they demonstrated how people evaluate information from different categories and how their dependence on intuition changes accordingly. In the following part some of the heuristics and the biases to which they lead are presented.

2.2.1 Representativeness

In the majority of the experiments conducted, people were presented with a question asking "What is the probability of the sample case A belonging to class B?" or "What is the probability of the sample case A stemming from process B?" (Tversky & Kahneman, 1972). People make judgments in terms of the extent to which A is representative of category B or in other words to what extent A resembles B. When A is highly representative of B, people think the probability of A belonging to the category of B is high. On the contrary if A does not resemble B, they judge the probability as being low. They gave the following example description in their paper:

"Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." (Tversky & Kahneman, 1972).

There is a list of possible occupations for Steve (e.g. engineer, pilot, librarian or secretary) and participants are expected to judge which one is more likely and which

one is least likely for Steve. In fact, they make probability estimations for each one of these occupations. In the representativeness heuristic, because Steve is a stereotypical member of the category librarian, people tend to judge the option librarian as most likely (Tversky & Kahneman, 1972). This explains the fact that people judge by similarity, not by probability (Tversky & Kahneman, 1973). In their work, Kahneman and Tversky (1972) examined the elements that determine representativeness of samples or events.

2.2.1.1 Similarity of sample to population: This determinant was explained with an example:

"All families of six children in a city were surveyed. In 72 families the exact order of births of boys and girls was G B G B B G. What is your estimate of the number of families surveyed in which the exact order of births was B G B B B B?"

A sign test was applied and the results were significant: 75 out of 92 people reported a mean of 30, meaning they thought the second order is less likely than the first one instead of giving the right answer of 36. In addition, they also performed an experiment with sophisticated psychologists as subjects (Tversky & Kahneman, 1971). The question asked is shown below:

"Suppose you have run an experiment on 20 Ss, and have obtained a significant result which confirms your theory ($x = 223$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 Ss. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group?"

The results were similar to the ones in the naive subjects' group. While the median should have been lower than .50, it was found as .85. According to Cohen (1962) these situations may cause a decrease in validity and reliability.

2.2.1.2 Reflection of Randomness

The second condition for event A to be representative of group B is for it to possess the features of its reference category. Capturing the randomness depends on the context but irregularity and local representativeness are important contributors (Tversky & Kahneman, 1971). For example;

"On each round of a game, 20 marbles are distributed at random among five children: Alan, Ben, Carl, Dan, and Ed. Consider the following distributions":

	First		Second
Alan	4	Alan	4
Ben	4	Ben	4
Carl	5	Carl	4
Dan	4	Dan	4
Ed	3	Ed	4

"In many rounds of the game, will there be more results of type I or type II?"

In this study people chose type I rather than type II because it was perceived as more random without including any pattern. Moreover, the reference category must be represented not only globally but also locally. There are many studies (Rapoport & Budescu, 1977; Kareev, 1995) demonstrating evidence for belief in local representativeness such as the gambler's fallacy. To give an example for this kind of fallacy; in a coin flipping game if the number of the heads are more frequent than the number of the tails, people think that there will be more tails than heads later on. The representativeness heuristic leads to many biases such as insensitivity to sample size, misconceptions of chance, misconceptions of regression, insensitivity to predictability and insensitivity to prior probability of outcomes which we will discuss in detail in the following chapter.

2.2.2 Biases Related to Representativeness Heuristic

2.2.2.1 Base-rate Neglect

Kahneman and Tversky (1973) explained this bias as making judgments relying on existing beliefs and neglecting the crucial prior probability information. Their first experiment examined this phenomenon in detail. In the first experiment, half of the subjects were told that there were 30 lawyers and 70 engineers in the population and the other half of them were told the opposite. Then they were presented with a description and they were asked to judge the probability of the person being a lawyer or an engineer. Clearly, participants in the first condition should have judged the probability that the description of a person being an engineer higher than that he/she is being a lawyer while the participants in the second condition should have given answers in the opposite direction. On the contrary, participants in both conditions made similar probability judgments without taking base-rates into consideration when the description was representative for each of the stereotypes.

2.2.2.2 Insensitivity to Sample Size

This bias reflects a tendency that people do not pay attention to the sample size when they make a probability judgment about a result obtained from a specific population drawn from this sample. The results of Kahneman and Tversky's (1972) study indicated that people assigned similar probabilities to the groups which are composed of 10, 100 and 1000 men in terms of average height being more than 6 feet. The judgments of probabilities should have been the highest for the most crowded group where there are 1000 men and for the group of 10 men, the probability should be smaller than the group of 100 men.

2.2.2.3 Misconceptions of Chance

In order for something to appear random, every small part of it is expected to represent the crucial features of the whole event. As in the famous coin and toss example, the sequence H-T-H-T-T-H is acknowledged as more random by subjects that the sequence of H-H-H-T-T-T and H-H-H-H-T-H (Kahneman & Tversky, 1972). In addition, there are other well-known examples such as the gambler's

expectation of a red after a long run of blacks and the trained psychologists' biased belief in "the law of small numbers". The former law is concerned with the bias where people tend to make estimations about a population relying on a small data set.

2.2.2.4 Insensitivity to Predictability

This bias was explained in terms of the factors that should play a role in the predictions that we make: reliability of the evidence and the expected accuracy of prediction. People basically rely on the description that they have been given and they do not take into account whether the evidence is reliable and the prediction is accurate. In other words, the normative statistical theory rules are not adapted when making predictions.

2.2.2.5 The Illusion of Validity

The illusion of validity indicates a tendency of feeling highly confident when making predictions if the descriptions are highly representative of their sample. People tend not to take into account other constraints on the accuracy of prediction. For example, psychologists are criticized in terms of their predictions raised from clinical interviews (Kahneman & Tversky, 1972).

2.2.2.6 Misconceptions of Regression

During the late 19th century Sir Francis Galton was the first one who mentioned regression to the mean or reversion to mediocrity as they called it in the past. It basically means that if one person scores higher than the mean in a test, he/she tends to get scores closer to the mean next time he/she is tested on an equivalent content. Humans' lack of attention to this phenomenon may result in harmful misinterpretations as it was in the flight training example of the Kahneman and Tversky (1983). In this example a pilot was faced with a compliment after a successful landing on the contrary he was criticised if the landing was not smooth enough. Because the people were not familiar with the concept of regression, they mistakenly thought that the reason why the pilot did better on his next flight after an unsuccessful landing was their criticism.

2.2.3 Availability Heuristic

The availability heuristic indicates the fact that people judge the likelihood of an event or frequency of a class depending on how easy it comes to his/her mind (Kahneman & Tversky, 1983). For example instances of larger classes are remembered better and faster than those of smaller classes. This heuristic is related with "the ease of relevant mental operation of retrieval, construction or association operations coming into mind" (Kahneman & Tversky, 1983, p.81).

2.2.3.1 Construction

Participants were given 3x3 matrices including nine letters for each problem and they were presented with two different types of tasks. The first one was an estimation task including 8 questions and the second one was a construction task.

For the estimation task people were given 7 seconds and they were supposed to estimate how many words with three or more letters they could extract from the matrix in 2 minutes. For the construction task, they were given two minutes to put the words that they constructed on a paper. While the mean was 11.9 for constructed words, the mean for the estimation of the words was 10.3.

2.2.3.2 Retrieval

The design of the experiment was similar to the construction experiment but the tests were different. In each problem there was a category like plants or English writers and within 7 seconds participants estimated how many examples they could find belonging to these categories in 2 minutes. For the other task, they wrote down the examples of the categories that they could think of in 2 minutes. Results showed that the mean of number of examples written was 11.7 and the mean of the estimation task was 10.8.

From these results, it can be inferred that availability can be quantified accurately by the numbers of items that were constructed and retrieved. The work of Bousfield and Sedgewick (1944) supported the idea that people are fast and accurate in their answers and they are able to make estimations without really constructing or retrieving any items. There are many biases that the availability heuristic causes phenomena such as illusory correlation, biases to the effectiveness of a search set, biases of imaginability and biases due to the retrievability of instances.

2.2.4 Biases Related to the Availability Heuristic

2.2.4.1 Biases Due to the Retrievability of Instances

A category or a set is perceived to be having more items than the other equally crowded one if the members of this category/set come to people's mind easier than the members of the other category. In an experiment there were two sets and each set comprised of equal numbers of women and men's names. In one of the sets there were more names for the famous men and in the other set there were more names for the famous women and people were asked to judge which set was more numerous. The results of two separate conditions showed that; they had a tendency to choose the one which contained more famous people's names even though the number of items in the two sets was equal.

2.2.4.2 Biases Due to the Effectiveness of a Search Set

In a study conducted by Galbraith and Underwood in 1973, participants were supposed to judge the frequency of concrete words and abstract words in English. It seems such a reasonable way to search for contexts where abstract and concrete words would appear. Consequently, people have a better chance to think of contexts where an abstract word appears than contexts where concrete words would appear.

For example; people probably think that the word "belief" would occur more frequently than the word "glass" in English. Availability of the contexts for abstract words caused people to judge that set as being more frequent than the set of concrete words.

2.2.4.3 Biases of Imaginability

When asked to judge the frequency of a class in which members have to be imagined, people tend to judge the frequency of that class depending on the imaginability of its members. Depending on this hypothesis, previous work showed that small groups tend to be judged as more crowded than larger groups. For example, when we are planning to do an activity such as diving, it is possible to picture all the things that can go wrong during the activity clearly in our minds. Consequently, we start to perceive this activity as more dangerous than it really is.

2.2.4.4 Illusory Correlation

This bias was first discovered by Chapman and Chapman in 1969. The rationale here is this: if two events are strongly associated with each other, people are likely to pair them more frequently even though there is no actual relationship between the two occurrences. Kunda (1999)'s work showed that people relied heavily on a few examples from their experiences of bad weather and pain occurring together.

2.2.5 Adjustment and Anchoring

These two heuristics refer to the situation where there is a number or a value obtained from a computation or a value indicated by the problem itself and people adjust it to give a final acceptable answer (Epley & Gilovich, 2005). It was first identified by Kahneman and Tversky in 1974 and Epley (2004)'s work made many contributions to this kind of heuristic. In certain experiments subjects were given an initial starting point by the experimenter and in others, subject determined it themselves. Adjustments were insufficient in both cases. For example, in a study conducted by Epley (2004) participants were asked questions such as: "When was the Declaration of Independence signed?" Then, they generated their own anchor by giving an answer that they knew was wrong but they also thought that it was close to the right answer such as "The Declaration of Independence was signed in 1776". The results revealed that people tried to stay close to their anchors, thus, they made inadequate adjustments.

2.3 Critical Perspectives on the Heuristics and Biases Approach

The Heuristics and Biases tradition helped researchers understand the flaws of human rationality but at the same time it was criticized sharply by many researchers in the field. Firstly, it was found limited in terms of the number of terms that it presented (Payne, 1980). Secondly, Gigerenzer and Murray (1987) said that even though bounded rationality theory formed the basis for this tradition, it deviated from its roots in an unreasonable way. Finally Lopes (1991) pointed out the considerations about the generalizability of the heuristics and biases approach. Mainly, it was found far too negative in terms of explaining the failures of human mind (Cohen, 1981; Einhorn and Hogarth, 1981).

The most argumentative claims against the Heuristics and Biases Approach are framed by Gerd Gigerenzer's work (Gigerenzer, 1991, 1994; Gigerenzer & Murray,

1987). In his critical articles, he basically argues that this program cannot explain how and why they work. In addition he claimed that it is not plausible for each heuristic to be applicable to more than one experiment. He also found heuristics ambiguous: Thus he doubted their acceptability as accounts of the biases of human reasoning. That is the reason why he suggested employing strategies and models to test the hypotheses and to examine the underlying mechanisms in the brain for inaccurate and accurate judgments (Gigerenzer, 1996).

2.4 An Overview of Three Dual-Process Theories of Reasoning

The development of dual-process theories is still continuing since the time of William James. They all basically say that human reasoning is based on two opposing systems. This topic has been studied by many researchers with differing descriptions of the two systems: Evans (2008) and Stanowich and West (2000) named them as System 1 and 2, Sloman (1996) called them associative and rule-based and finally Epstein and Kirkpatrick presented them as experiential and rational. In order to present a general overview, the most argued theorists, descriptions and comparisons are demonstrated. To begin with, Stanowich and West presented their Two-Systems Theory in their papers in 2000 and 2002 explicitly. They argued that the functional distinction between the two systems originates from the varying methods of reasoning that each individual has. In fact, they have conducted experiments with different reasoning tasks and found a strong correlation between these tasks and cognitive ability. According to Evans and Over (1996)'s Dual Process Theory, heuristics are not conscious and their task is to choose the appropriate representation that is linked to that problem area. Their System 1 is fast, associative and implicit. It depends on already existing knowledge and conscious awareness is not necessary for this system. On the other hand System 2 is slow, explicit and sequential. It has the ability to deal with logical problems. For their System 1 the crucial point is this: while its function is domain specific, its mechanism is domain general (Osman, 2004). Like Evans, Stanowich and West also claimed that System 1 depends on heuristics. They also made a distinction between System 1 and 2: System 1 is the evolutionarily rational one which is automatic, unconscious and context dependent; on the other hand, System 2 is the analytical and instrumentally rational one that has the ability to take control and make abstractions. Unlike Evan (1984)'s theory, Sloman (1996) did not make a distinction between the System 1 and System 2 depending on conscious awareness. In fact, he claimed that the difference between the two systems comes from their differing underlying computational mechanisms. Moreover, according to his theory humans are only aware of the output not how it was created in System 1. In System 2 they are consciously aware of the process and the output (Osman, 2004). Sloman (2002) defines his two systems as "interactive" and he made a distinction between them in terms of diagnostic tasks that are in rapport with a Criterion "S" as he called it. In an evolutionary sense, Stanowich and West (2000) distinguished these two systems in terms of the success at the gene level and at the singular organism level (Osman, 2004). Over and Evans (1997) evaluated System 1 as containing all accomplishments through the past which had positive contributions to survival. In addition they argued that humans need rational System 2 which appeared to join the

first system later through the evolutionary process. In a comprehensive overview of the literature for the theories explaining base-rate neglect Sloman and Barbey (1997) argued that human judgment should be studied in terms of dual processing approaches. In their opinion, while one system has the ability to do reasoning with links between varying set representations and with natural frequencies, the other system accounts for systematic errors. Basically, this view supports the idea that presenting base-rate problems by using natural frequencies rather than presenting them in the form of single-event probability will increase the usage of base-rate information. In their rationale the rules that people are able to reach and apply when they try to solve base-rate problems are in accordance with natural frequencies. As stated earlier, people need a rule-based system in order to solve problems according to the Bayesian perspective. Contributing to this, natural frequency views develop arguments in relation with rule-based processes but they do not study the relation between associative processes in Bayesian inference (Sloman & Barbey, 1997). In a recent study Evans and Stanovich (2013, p. 225) listed the characteristics of dual process theories of higher congition as follows:

Table 1 Clusters of Attributes Frequently Associated with Dual-Process and Dual-System Theories of Higher Cognition

Type 1 process (intuitive)	Type 2 process (reflective)
Defining features	
Does not require working memory	*Requires working memory*
Autonomous	*Cognitive decoupling; mental simulation*
Typical correlates	
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Biased responses	Normative responses
Contextualized	Abstract
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability
System 1 (old mind)	System 2 (new mind)
Evolved early	Evolved late
Similar to animal cognition	Distinctively human
Implicit knowledge	Explicit knowledge
Basic emotions	Complex emotions

2.5 Preceding Studies of Base-Rate Facilitation

One of the earliest and highly influential base-rate studies belongs to Kahneman and Tversky (1973). Here is the instruction given to the participants:

"A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100."

They were also informed that a group of experts participated in this study and their results revealed high accuracy rates. Moreover, subjects were told that they were going to get bonus payment according to the similarity levels of their results to the experts'.

There were two experimental conditions: in the first one, participants were instructed that there were 30 engineers and 70 lawyers in the set, on the contrary the second group were informed that there were 70 engineers and 30 lawyers. Then each subject saw five descriptions. For example:

"Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles."

Then they judge the probability of Jack being one of the 30 engineers in the set by indicating the percentage out of one hundred. After that, subjects were not given any personality descriptions and they were asked to decide the probability of the person being one of the 30 engineers of 100 people out of one hundred percent (Kahneman & Tversky, 1973).

The result of this study demonstrated that people did not attend to base-rate information when they were given personality descriptions but, they were more successful at estimating the probabilities when there was no other information to take into account. Moreover, making changes in the amount of base-rates caused a little – though insignificant – difference between the estimated.

The improvements reported in the literature of base-rate neglect or base-rate fallacy studies revealed that base-rates are not totally neglected (Bar-Hillel, 1983). In fact, they are used in varying degrees in terms of problem format and style of representation (Koehler, 1996, p.6). Koehler's Base-rate Usage Model captures the various factors influencing reasoning in base-rate problems:

According to this model, subjects tend to attend more to the information that varies across trials (Fischhoff *et al.*, 1979; Schwarz *et al*., 1991). One plausible explanation might be this: participants use these base-rates as a cue to problem solving. Effect of direct experience was demonstrated in a study conducted by Manis *et al*. (1980) in which participants received feedback after each trial. It is also important for samples to be well-defined. For example, if we have a look at the cab problem of Tversky and Kahneman (1980) below, we can see how ambiguous its sample space was:

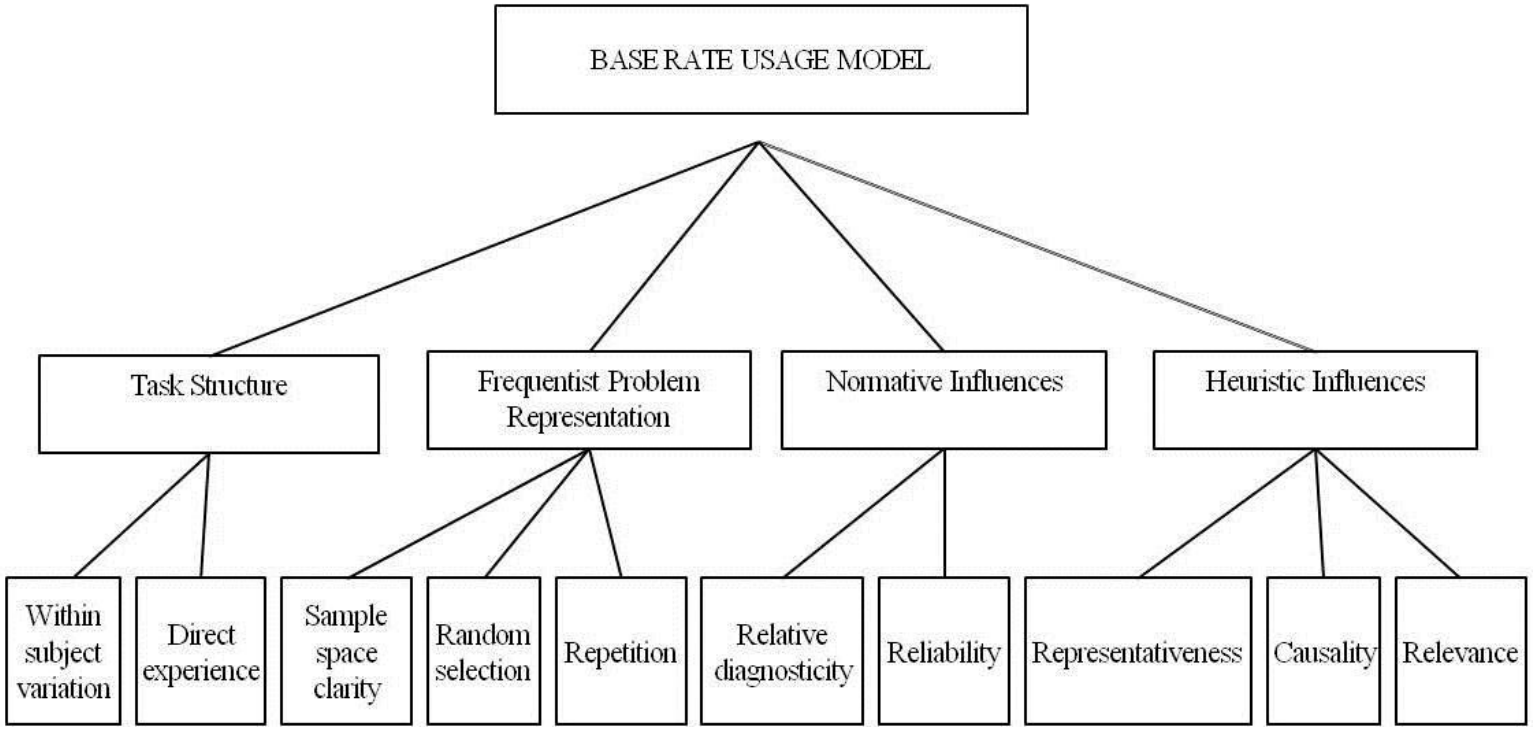


Figure 1 Base rate usage as a function of task structure, problem representation, normative influences and heuristic influences (Koehler, 1996, p.6)

"A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data: (i) 85% of the cabs in the city are Green and 15% are Blue, (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases. Question: What is the probability that the cab involved in the accident was Blue rather than Green?"

The randomness factor is crucial also. In an experiment conducted by Gigerenzer *et al*. (1988) they found out that people performed better when they actually participated in the random selection process by themselves rather than other participants who were told that the descriptions in the studies were selected randomly. Moreover, repetitive sampling from the same reliable and well-defined population has a posisitive effect on participants in terms of base-rate use.

Previous studies revealed that if base-rate problems are represented in terms of frequencies rather than probabilities, use of base-rates increases (Griffin & Buehler 1999). Barbey & Sloman (2007) suggested an explanation for this finding: effect and sample sizes that are represented by natural frequencies have such characteristics as emphasizing the set structure of the problem. This crucial feature of the set structure makes it easier to solve the base-rate problems. Using different base-rates in different trials (Bar-Hillel & Fischhoff, 1981), giving the base-rate information after the descriptions (Krosnick, Li & Lehman, 1990), doing the sampling of the descriptions randomly every time (Gigerenzer, Hell & Blank, 1988) and instructing subjects to think like a statistician (Schwarz, Strack, Hilton, & Naderer, 1991) all help overcoming the base-rate neglect. Kohler (1996) also claimed that when participants engage in a task that enables them to learn base-rates without conscious awareness, base-rate utilization occurs. Moreover they are used more frequently when they are perceived as more reliable and when they reveal more characteristic information than given descriptions. Evans *et al.* (2002, Experiment 5) found out that use of real-world beliefs instead of actual probability statements for prior probabilities, caused people to make better judgments. Lastly, participants who have greater working memory capacity and have relatively higher intelligence are more likely to rely on analytical thinking and therefore are more likely to be successful than other people in base-rate problems (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2013; Pennycook, Cheyne, Seli, Koehler, & Fugelsang, 2012; Stanovich & West, 2008).

Although a lot is known already about human reasoning, based on these behavioral experiments in the tradition of the Biases and Heuristics and Dual Process Theories, relatively little is known about which on-line cognitive processes may support these higher cognitive processes. The lack of studies using eye-tracking method in order to study base-rate neglect encouraged this thesis to investigate how reasoning processes under various conditions manifest themselves differently in on-line measures of perceptual processing as revealed by eye-tracking methodology. This relation might look like, e.g., subjects' preferred looking to certain regions on the

screen (text or graphic) will reveal on which kind of information they base their reasoning. Eye-tracking may provide important additional and quantifyable information as to these underlying cognitive processes. Based on the findings in the literature, the present thesis studies the effect of using graphics and frequencies as presentation format, random sampling, and various amounts of feedback on solving base-rate problems. Off-line and on-line measures were collected in the form of accuracy scores, response times, and several eye-tracking measures. The latter are supposed to reveal how higher cognitive reasoning processes might be supported by on-line perceptual and attentional processes.

CHAPTER 3

METHOD

A pilot study and a main study were conducted within the framework of this thesis with the aim of finding out to what extent presenting base-rate problems with different modalities (visual, verbal) has an effect on base-rate neglect. Subsequent to creating a problem pool to select the most suitable items and conducting a pilot study, participants' performance in a base-rate neglect reasoning task, including tracking of their eye movements was assessed.

3.1 Pilot Study

3.1.1 Participants

A total of 86 female and 86 male undergraduate and graduate subjects (52 of them rated base-rate problems and 120 of them participated in the actual pilot experiment) majoring in different departments from Middle East Technical University (METU) participated in this study. They were recruited from the entrance of the university's library. Participants were native speakers of Turkish and their age ranged from 18 to 32. They filled out the participant form and a consent form before starting the study. Subjects also reported that they were not on any kind of drugs that may affect their cognitive abilities.

3.1.2 Apparatus and Stimuli

3.1.2.1 Base-rate Problem Rating Task

In order to carry out the pilot study, a 36-item base-rate problem pool was generated. There were three categories that represent these problems; No-conflict, Conflict and Neutral. Each category included 12 items. In the No-Conflict category, the stereotypical information conveyed in the description of a person belonged to a larger majority group in the population. In contrast when the description represented the characteristics of a smaller minority group in the population, a conflict between the good fit of the description with the minority group and its low base-rate occurred. In the Neutral category, there were no hints that if the protagonists described in the problems belonged to either the bigger majority or the smaller minority in the population. Eleven out of twelve base-rate problems were translated into Turkish from the study of De Neys and Franssens (2009) and then they were adapted for the Turkish sample. Originally, De Neys and Franssens used Kahneman and Tversky (1973)'s problems in their study. In the current study, their third congruent problem was not used. In addition, twenty four more base-rate problems were created based on inoffensive stereotypes including race,

age, gender etc. In each of these thirty six problems a person was described (without any base-rate information given) and a one-sentence statement about the people in the descriptions were printed on a paper side by side with a 5-point Likert-type scales (1 indicating "totally disagree" and 5 indicating "totally agree"). The statements in the no-conflict condition were in the same direction with the description, the statements in the conflict condition were in the opposite direction with the description and lastly the statements in the neutral group did not have any direction at all. Fifty two subjects rated the problems from 1 to 5 by answering how likely the statements were true for each description. These following examples are from these three categories (All items can be seen in Appendix A).

No-Conflict:

1. Aysegul is 35 years old. She writes critiques for a magazine. Her husband works in a university. Aysegul likes painting and taking photographs.

How likely do you think the information given below about Aysegul is true?

Aysegul likes to watch National Geographic Channel.

Conflict:

15. Gokce likes to go dancing with her friends at the weekends. Generally, she prefers fast food and she has a little piercing in her belly.

How likely do you think the information given below about Gokce is true?

Gokce is fifty years old.

Neutral:

26. Toprak is 19 years old. He studies in Istanbul and he doesn't have a girlfriend. He bought a second-hand car with the money he saved.

How likely do you think the information given below about Toprak is true?

Toprak plays drum.

The first twelve items were no-conflict problems, the next twelve items were conflict problems and the last twelve items were neutral problems. Every subject rated the items in the same order on a separate answer sheet. SPSS Statistics 17.0 was used to calculate the means and the standard deviations. Items for the final stimulus set were selected according to the means: no-conflict items should display the highest agreement; conflict items the lowest; and neutral items should be in the middle range. For the no-conflict condition items 1, 3, 5, 6, 7 and 11 were selected because they had the highest mean scores. In contrast, for the conflict condition problems with the lowest six mean scores were selected: 13, 15, 17, 20, 21 and 22. Moreover for the neutral condition, items 26, 29, 31, 32, 33 and 35 were selected because they had the means that were closest to 3. (see Table 2).

Table 2 The Descriptive Statistics for the thirty six items in the problem pool

Descriptive Statistics

	N	Mean	Std. Deviation
One	52	3,8846	,83205
Two	52	3,2885	1,10855
Three	52	3,6346	1,10309
Four	52	2,7885	,99679
Five	52	4,5000	,75407
Six	52	3,8654	1,17204
Seven	52	3,4808	,87426
Eight	52	2,4423	1,01775
Nine	52	2,8077	,97092
Ten	52	2,8846	1,06004
Eleven	52	3,7115	,99679
Twelve	52	3,0962	1,12476
Thirteen	52	1,9231	,96703
Fourteen	52	2,4231	,91493
Fifteen	52	1,6154	,77089
Sixteen	52	2,3654	,90811
Seventeen	52	1,9423	1,07400
Eighteen	52	2,9808	,98000
Nineteen	52	2,0769	,98710
Twenty	52	1,7885	,97692
Twentyone	52	1,9615	1,02826
Twentytwo	52	1,4231	,75006
Twentythree	52	2,3269	1,04264
Twentyfour	52	2,9615	,94892
Twentyfive	52	3,4615	,82751
Twentysix	52	2,9808	,99981
Twentyseven	52	2,6538	1,02679
Twentyeight	52	2,7885	,87080
Twentynine	52	3,0000	,88561
Thirty	51	3,0392	,74728
Thirtyone	52	3,0385	,76598
Thirtytwo	52	3,0385	,65564

Table 2 (continued)

Thirtythree	52	3,3077	,80534
Thirtyfour	52	2,9038	,77357
Thirtyfive	52	3,0385	,86232
Thirtysix	52	2,0769	1,15209
Valid N (listwise)	51		

In order to find out if these three conditions were significantly different from each other or not, a Univariate Analysis of Variance was conducted. In this item-wise analysis the average rating scores of the 18 problems across subjects were the dependent variable while the condition (no-conflict, conflict and neutral) was taken as the independent variable. Results demonstrated that the difference between three conditions was statistically significant *($F_{(2, 15)}$ = 111.597, p < .001, η_p^2 = .937)* (see Table 3).

Table 3 Pairwise Comparisons of the three conditions

Dependent Variable:Average

(I) Category	(J) Category	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
NC	C	,723	,035	,000	,629	,816
	N	,255	,035	,000	,161	,349
C	NC	-,723	,035	,000	-,816	-,629
	N	-,467	,035	,000	-,561	-,374
N	NC	-,255	,035	,000	-,349	-,161
	C	,467	,035	,000	,374	,561

No-conflict-items showed significantly higher levels of agreement (M=3.84, SD=.35) than neutral items (M=3, SD=.05) and neutral items showed significantly higher levels of agreement than conflict items (M=1.77, SD=.21).

3.1.2.2 Pilot Reasoning Task

The purpose of the pilot study was to investigate the effects of different modalities (verbal, graphical) on reasoning about base-rate problems. In particular, it aimed to find out if a graphical presentation of the crucial probability information, prior to or simultaneously with the verbal problem statement, improves proper reasoning about base-rate problems in comparison to the classical verbal-only presentation mode. For this purpose pie-charts were created showing the base rate of the minority group in the

overall population. There are four conditions in the pilot study: (1) Pie-chart Condition, (2) Verbal Statement-only Condition (base condition), (3) Numeric Percentage-only Condition and (4) Irrelevant Picture Condition. Accepting the Verbal Statement-only condition as base condition, Numeric Percentage-only condition and Irrelevant Picture condition were created to test the hypothesis that presenting any visual cue for base-rate problems would not help solving the problems. Moreover, Numeric Percentage-only condition was used to test the hypothesis that presenting the crucial probability information graphically would help participants more in order to make accurate judgments than they were presented the crucial probability information verbally. The pie-charts in the first condition were created equal in size with Microsoft Office Excel 2007. All verbal statements were written in Times New Roman with a font size of twelve. For the fourth condition the most representative images for the base-rate problems were selected through some sites on the internet. For example; if there were two different channels in the problem, their logos were used and if the problem was asking whether someone is a basketball player or a billiard player, a highly representative pictures of these two sports were used (for all the pictures used in the experiment see Appendix B). They were prepared in Adobe Photoshop CS. Cool Record Edit Pro with the same resolution which was 130x160 dpi. Considering the fact that De Neys and Glumicic (2008) did not find any performance difference between slightly different base-rates, for this study base-rates of 995/5, 996/4 and 997/3 of a population of 1000 individuals were used in the problems and also in the making of the pie-charts. The reason for choosing these probabilities was making the differences as extreme as possible. Since only 18 base-rate problems were chosen from the pool, each condition (no-conflict, conflict and neutral) had 6 problems. All three above-mentioned base rates couples were used twice in one condition and were assigned to the problems randomly by hand. E-Prime 2.0 was used to conduct the experiments on a Lenovo Ideapad Z370 laptop with operating system Windows 7. Participants were seated in front of the laptop that was located on a table in a private study room METU's university library. There were 30 participants (15 males and 15 females) in each of the four experimental conditions. Subjects did not get any verbal instructions before any of the experiments; the instructions were given within E-Prime. The instructions gave them an idea about the experiment and which keys to press when. The locations of correct and wrong answers were counterbalanced. For the half of the problems the correct answers were located on the top (1) and for the other half of the experiments they were located at the bottom (2). At the beginning of the first condition (graphical, verbal) participants read the instruction screen. Then, the moment they were ready start the experiment they pressed "space". On the next screen they saw a pie chart displaying the distribution of the two sub-populations in the problem (see Figure 2).

Figure 2 The first stimulus used in the Pie-chart Condition

Subjects pressed "space" when they were ready to proceed to the next screen. Then, the problem and the same but smaller sized pie-chart they had seen in the previous screen were presented together (see Figure 3).

Figure 3 The second stimulus used in the Pie-chart Condition

Then they had to make the judgment by pressing the keys “1” or “2”.

The location of the pie-charts and the problems were chosen randomly by the program. Participants sometimes saw the pie-charts on the left side and the problems on the right side and sometime it was the opposite case. The second condition was composed of only verbal statements of the same 18 problems (see Figure 4). The same keys on the keyboard were used to answer the questions.

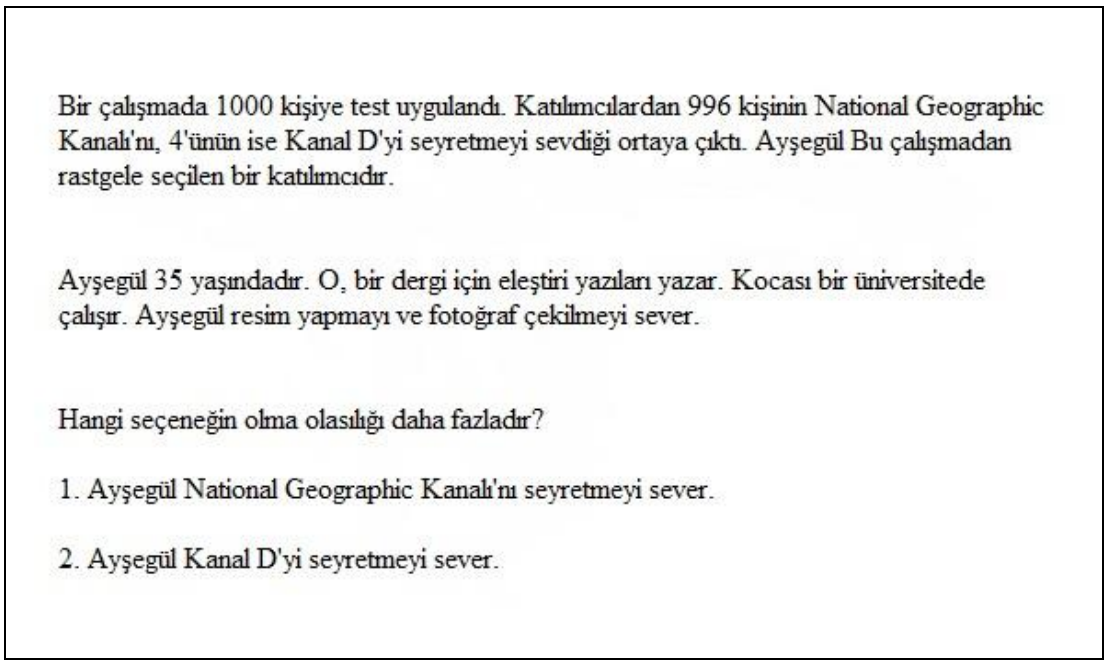


Figure 4 An example stimulus used in the Verbal Statement-only Condition

The third condition’s design was similar to the first conditions’ except that instead of pie-charts, there were percentages written on the screen demonstrating the distribution of the population in the problems (see Figure 5 and Figure 6). The same keys were used in this condition.

In the fourth and last condition participants saw only one screen which included the problem on one side and two pictures related to the problem domain (but giving no clue to the base-rate information) on the other side. Locations (left or right) of the verbal statements and the pictures were chosen randomly by the program (see Figure 7).

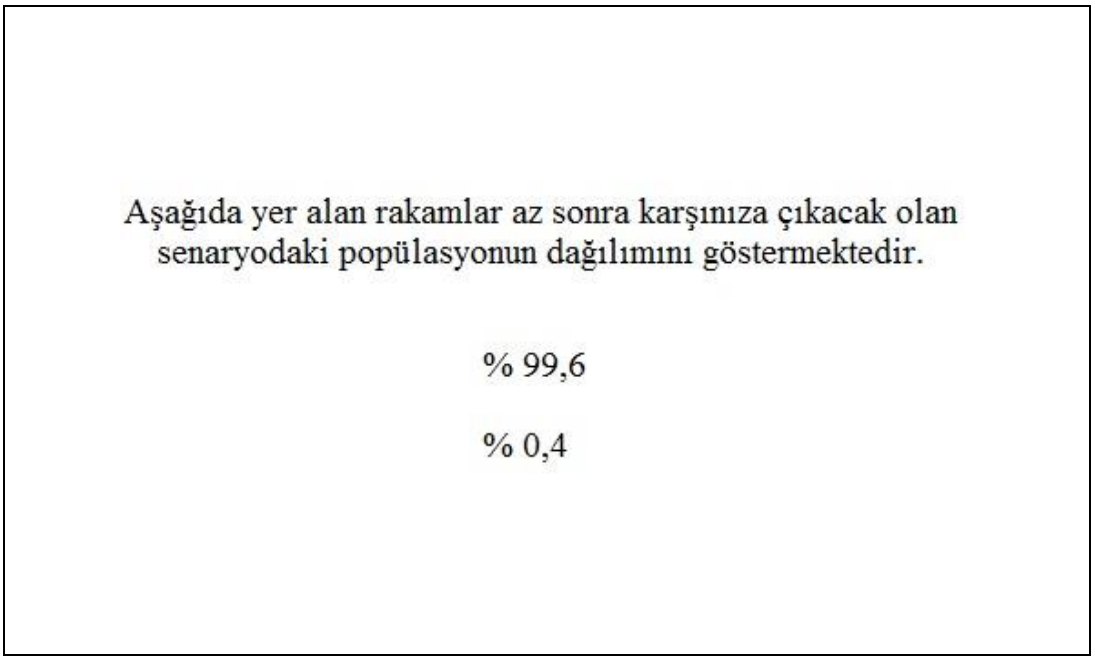


Figure 5 The first stimulus used in the Numeric Percentage-only Condition

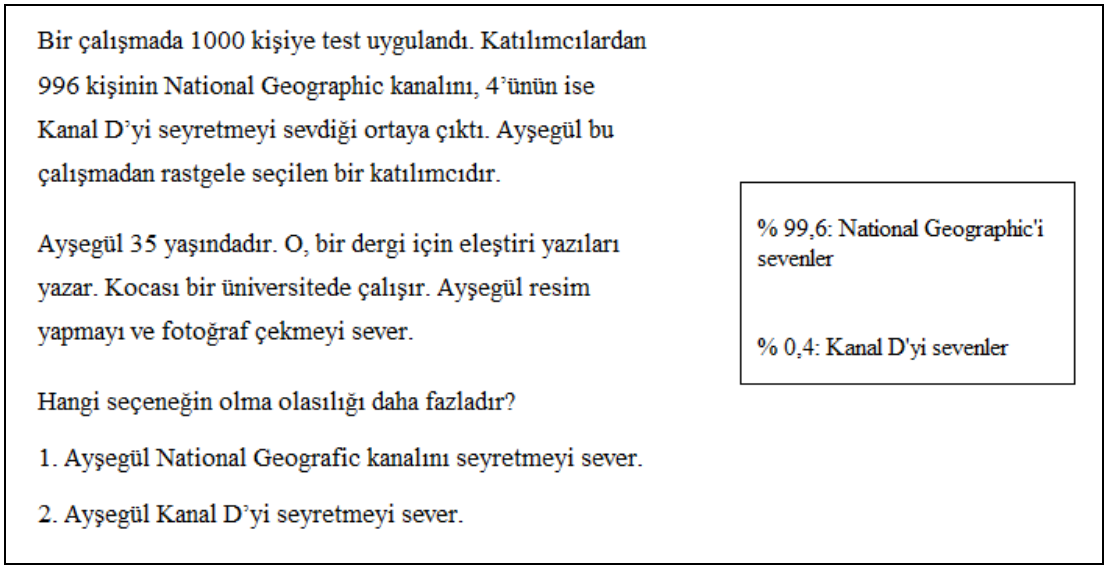


Figure 6 The second stimulus used in the Numeric Percentage-only Condition

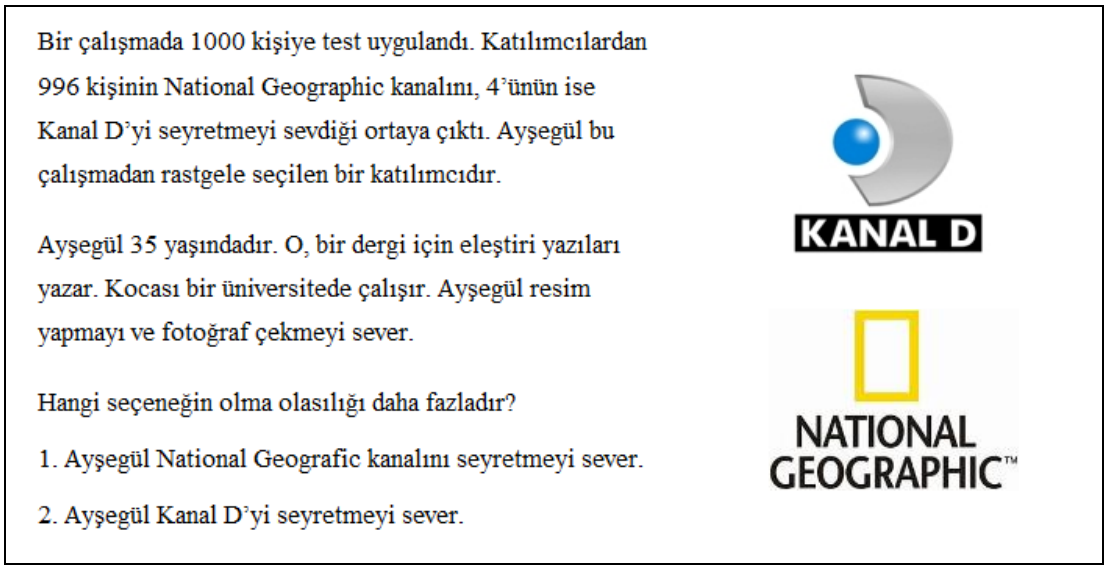


Figure 7 An example stimulus used in the Irrelevant Picture Condition

A 2x4x3x2 mixed analysis of variance (ANOVA) was applied. Gender was included in the analysis as a control variable and Part was added to see whether any learning occurred. Gender (Female/Male) and Experimental Group (First/Second/Third/Fourth) were taken as between subjects factors while Condition (No-conflict/Conflict/Neutral) and Part (First/Second) were taken as within subjects factors.

The dependent variable was accuracy rates translated into percentages. The right answer was the one that represents majority group. Although the results revealed a main effect of Condition *($F_{(2,\ 224)}$ = 463.586, p < .001, η_p^2 = .805)*, there was no statistically significant difference between these four experimental groups. In terms of the main effect of condition, the no-conflict condition *(M =0.8982, $S.D.$ = 0.13024)* had higher accuracy rates than neutral condition *(M =0.63, $S.D.$ = 0.26)* and conflict condition *(M =0.14, $S.D.$ = 0.22).*

3.2 Main Study

3.2.1 Participants

A total of 36 (18 female and 18 male) undergraduate and graduate subjects studying in different departments from METU participated in the main study. They were recruited from the entrance of the Human Computer Interaction (HCI) Laboratory located inside the Computer Center of METU. Participants were native speakers of Turkish and their age ranged from 18 to 32. They filled out the participant form and a consent form before the start of the experiment. Subjects also reported that they were not on any kind of drugs that might affect their cognitive abilities.

3.2.2 Apparatus and Stimuli

The same eighteen base-rate problems selected for the pilot study were used again in the main study. E-Prime 2.0 was used to present the stimuli and to measure accuracy rates and reaction times and Tobii Studio T120 was used to measure total fixation duration, total visit duration and visit count of the areas of interest (verbal, graphical). E-Prime was executed on Lenovo Ideapad Z370 laptop and Teamviewer 9 was used to connect the laptop and the test computer to whom the Tobii Studio T120 was attached. In these experiments different base-rate levels namely 9/1, 8/2 and 7/3 (comprising a total sample of 10 individuals) were chosen to be able to present these levels as frequencies and to make the visualization of distributions easier for participants. For the main study instead of the extreme probability rates, the values which had higher ambiguity were chosen in order for participants to pay more attention to the changing prior probabilities. Eighteen base-rate problems 9 (3 from each condition) for the training phase and 9 for the test phase were assigned randomly. There were three types of stimuli that were used in all the three experimental conditions. The first stimulus was composed of ten neutral faces (2 centimeters in diameter) located on a white background with equal distances in between from which participants had to choose one randomly. The next stimulus was composed of 10 white and light grey circles (3.8 centimeters in diameter) with equal distances in between representing individuals from the majority and minority groups, respectively. In addition to these circles, there was a sentence written in Courier New with the font size of 16 saying "The distribution of the individuals in the study". On the final screen, the base-rates problems were written in Courier New with the font size of 8 next to the frequency distribution of the population that they had seen in the previous screen and this time with a little light grey and a little white circle under the graphic together with the text stating what they represented. The location of the problem and the graphic (being on the left or being on the right) were chosen by E-Prime randomly.

Some participants saw the base-rate problem on the left side of the screen while the graphic being on the right side of the screen, other participants saw them in opposite locations. The locations of the correct and wrong answers were counterbalanced. For half of the problems the correct answer was "1" and for the other half of the problems the correct answer was "2".

3.2.3 Procedure

Participants were seated in front of the test computer which had Tobii Studio in the HCI Lab. First of all, they follow a red circle's movements on the screen in order for the program to do the calibration. After successful calibration (60% or higher) they were given brief information without any clue on how to solve the problems. There were three experimental conditions: (1) No Feedback condition, (2) Color Feedback Condition and (3) Double Feedback Condition. there were two phases, one in which they performed a random choice of one individual of the sample and then had to consider the likelihood of that individual to belong to the minority or majority group of the overall sample a subsequent phase in which they performed the base-rate problem. The idea was that if participants understood that their randomly chosen individual was most likely to belong to the majority group they would perform better on the base-rate problem, resisting the

temptation to attribute the highly representative characterization of the minority group to that individual. In phase (1) participants were basically told that at first, they were going to see faces on the screen and these faces represented people in our study. It was said that, they could choose which ever they wanted by using the keys on the keyboard from 1 to 10 (see Figure 8). It was explained that the face located on the very left side on the top was number 1 and the face located on the very right side at the bottom of the screen was number 10.

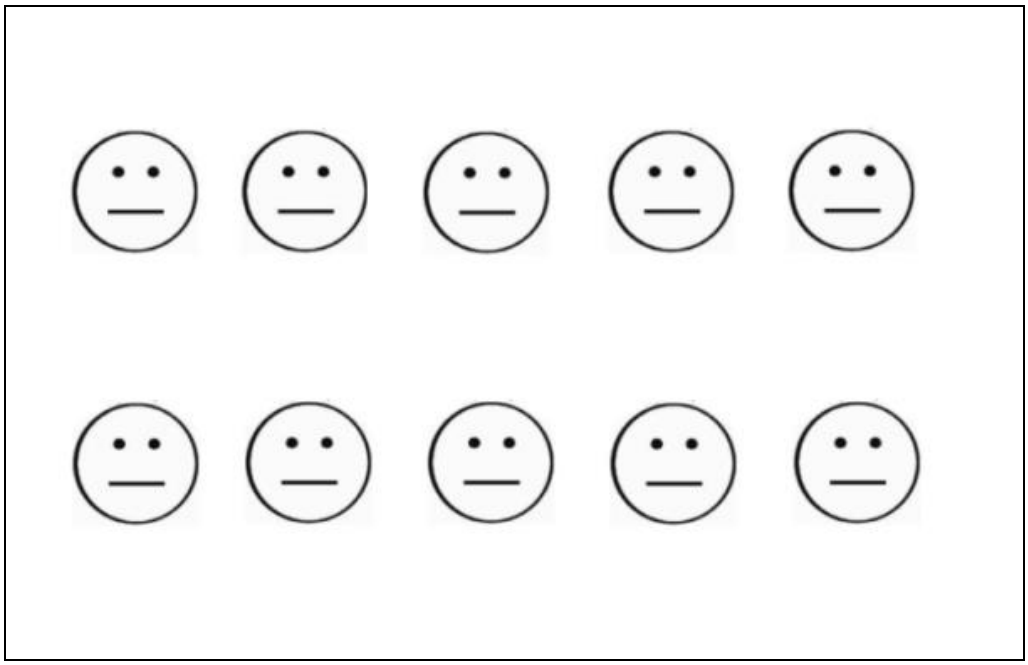


Figure 8 The stimulus representing the population in the problem space

After they had chosen one person, they were told that they were going to see a distribution of the population that the person they had chosen previously belonged to.

They were asked "Which group do you think the person you chose is most likely to belong to? If you think he/she is most likely to belong to the white group please press "b" on the keyboard but if you think he/she is most likely to belong to the grey group please press "m" on the keyboard (see Figure 9). Before the experiments, the participants in three experimental groups were instructed that the faces in the random sampling process and the circles were not in the same order. Thus, we tried to avoid that participants might identify the individual that they had chosen randomly with the circle at that place on the next slide which belonged to either the minority or majority group, by chance.

However, participants were not given any feedback on whether their choice was correct or incorrect. In phase (2), they were presented the base-rate problem side by side with the same frequency distribution of the population except that this time it was given together with information what the grey group and the white group means, under the graphic display. They were asked to choose either one of the answers. "If you think the

first answer is correct, please press 1 on the keyboard but if you think the second answer is correct, please press 2 on the keyboard" (see Figure 10).

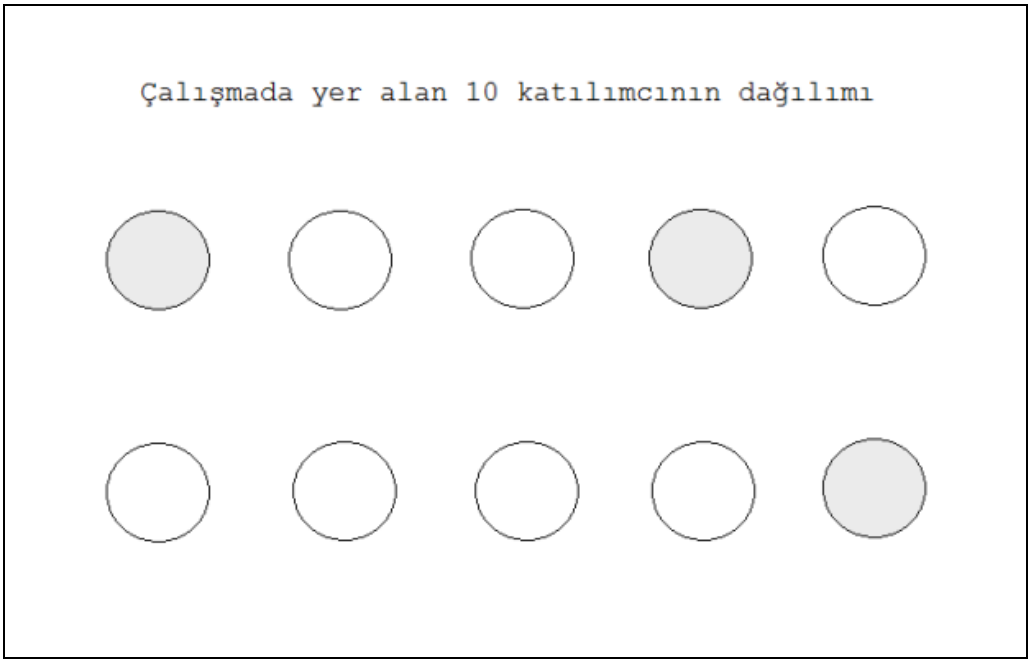


Figure 9 The stimulus demonstrating the frequency distribution of the population in the problems

After pressing either "1" or "2", participants proceeded to the next problem had the same structure as explained above. Each participant solved 18 problems (6 of either condition: (1) no-conflict, (2) conflict, (3) neutral).

The second experimental group (Color Feedback Group) also underwent two experimental phases. In phase (1) they performed a training phase before the actual experiment. In phase (1) they were told that they had to answer a few questions and feedback would be given to their answers, in addition to the instructions for the No Feedback Group. First, they saw the screen with the ten faces in order to choose someone and then the frequency distribution screen appeared. When they answered to the question "Which group do you think the person you chose is most likely to belong to?" they were given a feedback. The feedback was either "Dogru!"(Right) or "Yanlıs!" (Wrong). The correct answer was always the one in which the majority group was chosen as the most likely group to which their randomly chosen individual belonged. They were presented 8 couples of these questions for training. Participants who gave more than two wrong answers were excluded from the study during the analyses. 3 out of 15 participants were excluded from subsequent analyses for that reason. In phase (2) participants solved the same 18 base-rate problems for the rest of the experiment, however and they did not get any feedback on the correctness of their answers.

The last experimental group was the Double Feedback Group. The procedure in this group consisted of a training phase and a subsequent test phase. Training as well as test

phase consisted of two phases again. In this group, in phase (1) of the training phase, participants performed the random choice and answered the probability question to which group the chosen individual would most likely belong to. In phase (2) they were given the base-rate problem.

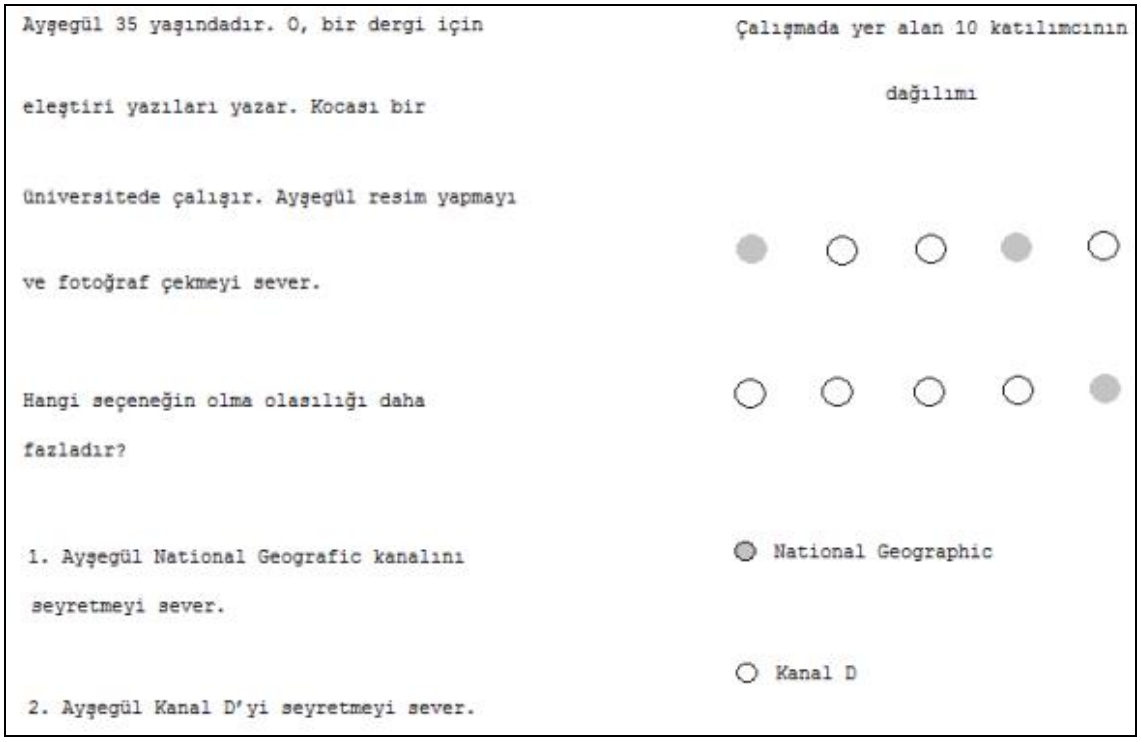


Figure 10 The stimulus comprised base-rate problem and the frequency distribution of the problem

Crucially, they received double feedback on the correctness of their answers in both phases throughout the training phase which consisted of (9 problems: 3 items were from the no-conflict condition, 3 items were from the conflict condition and 3 items were from the neutral condition). Participants were given feedback for both the color questions in phase (1) and for the base-rate problems in phase (2) during the training phase, but they did not receive any feedbacks for the second half of the experiment, the test phase, which consisted of 9 base-rate items as well.

CHAPTER 4

RESULTS

Both item- and subject-wise analysis were run for cross-checking effects in both analyses and in order to know whether any effect goes back to the particular variance in terms of differences within the test items or in terms of differences within the participants. Since no major differences were found between item- and subject-wise analyses we dropped subject-wise analysis.

4.1 E-Prime Analyses

Accuracy scores (sum scores of correctly answered items) of judgment and reaction times (RTs in seconds) of No Feedback Group, Color Feedback Group and Double Feedback Group were compared in terms of no-conflict base-rate problems, conflict base-rate problems and neutral base-rate problems.

4.1.1 Accuracy Scores

A 3x3 mixed ANOVA (condition: no-conflict, conflict, neutral; group: no feedback, color feedback, double feedback) for the accuracy scores of the scenario questions (in each group, there were three conditions; in each condition there were six color questions and six scenario questions) was conducted with item-wise arranged data. In the analysis condition was a between-subject factor whereas group was a within-subject factor. The results revealed a main effect of group *($F_{(2, 30)}$ = 13.75, $p < .001$, $\eta_p^2 = .478$)* as well as of condition *($F_{(2, 15)}$ = 17.34, $p < .001$, $\eta_p^2 = .69$)*. Accuracy scores of scenario questions in the No Feedback Group (*M* =5.77, *SE* =.47) are lower than the Color Feedback Group (*M* =7.33, *SE* =.45) and Double Feedback Group (*M* =7.61, *SE* =.38). Participants were less accurate in their judgments for the conflict problems (*M* =3.72, *SE* =.66) then they were for the neutral (*M* =8.44, *SE* =.66) and no-conflict (*M* =8.55, *SE* =.66) problems.

A 3x3 mixed ANOVA (condition: no-conflict, conflict, neutral; group: no feedback, color feedback, double feedback) for the accuracy scores of the color questions (in each group, there were eighteen items) was conducted with item-wise arranged data. Condition was a between-subject factor whereas group was a within-subject factor. The results revealed a main effect of group *($F_{(2, 30)}$ = 10.53, $p < .001$, $\eta_p^2 = .41$)*. In the Color Feedback Group (*M* =10.88, *SE* =.21) participants were more successful in their predictions for the color questions then the people in the No Feedback Group (*M* =9.11, *SE* =.26) and Double Feedback Group (*M* =8.83, *SE* =.58).

Lastly, a repeated measures ANOVA was conducted to find out whether the training and the test parts of the Double Feedback Group differed significantly in terms of the

accuracy scores (there were nine items both in the training part and in the test part) of the scenario questions in different conditions (no-conflict, conflict, neutral). Part (training, test) was taken as a between-subject factor, whereas condition (no-conflict, conflict, neutral) was taken as a within-subject factor. Results revealed a significant main effect of condition *($F_{(2, 44)}$ = 4.97, p < .05, η_p^2 = .18)* and part *($F_{(1, 22)}$ = 4.87, p < .05, η_p^2 = .18)*. In the Double Feedback Group participants were more successful in their answers for the no-conflict problems (*M* =2.12, *SE* =.14) and neutral problems (*M* =2.12, *SE* =.19) than for the conflict problems (*M* =1.45, *SE* =.21). Moreover, accuracy scores were higher in the test part (*M* =2.16, *SE* =.16) than the training part (*M* =1.63, *SE* =.16).

4.1.2 Reaction Times

A 3x3 mixed ANOVA with "group" (No Feedback; Color Feedback; Double Feedback) as between-subject factor and "condition" (No-conflict; Conflict; Neutral) was run to investigate the reaction times of item-wise arranged data. The results revealed a main effect of condition *($F_{(2, 15)}$ = 8.13, p < .005, η_p^2 = .63)*. Participants made judgments for the items in the neutral condition (*M* = 19.45s, *SE* = 588.73) more slowly than they did in the no-conflict condition (*M* = 15.85, *SE* = 588.73) and conflict condition (*M* = 15.69, *SE* = 588.73).

4.2 Tobii Studio Analyses

In the experiments the screen was portioned in two areas of interest (graphic, text).

4.2.1 Sign Test

A sign test was conducted with the data where all conditions (no-conflict, conflict, neutral) and groups (no feedback, color feedback, double feedback) were combined to see if there was a significant difference between the two areas of interests (graphic, text) in terms of first looks and last looks overall. The results demonstrated a significant difference between the number of first looks (Z =2.97 p = .003) and last looks (Z =3 p = .003) to the graphic and text areas. The number of first looks in the text area (*M* = 10.25, *SD* = 2.02) were higher than the number of first looks in the graphic area (*M* = 7.64, *SD* = 1.82). Similarly the number of last looks in the text area (*M* = 11.25, *SD* = 3.8) were higher than the number of last looks in the graphic area (*M* = 6.64, *SD* = 3.85).

A second sign test was conducted with the data where all conditions (no-conflict, conflict, neutral) were combined to see if there was a significant difference between the three experimental groups in terms of their first looks and their last looks for both areas of interests (graphic, text). A binomial sign test revealed a significant difference (p < .05) between the number of first looks to the graphic area and the text area in the Double Feedback Group as compared to the other groups. In addition in the same experimental group a significant difference (p < .001) between the number of last looks to the graphic area and the text area was found. The subjects looked at the text area (*M* =10.42, *SD* =1.5) firstly more frequently than they looked at the graphic area (*M* =7.58, *SD* =1.5). Similarly, the number of times the participants looked at the text area (*M* =14.08, *SD* =2.46) lastly are higher than they looked at the graphic area (*M* =3.92, *SD* =2.46).

4.2.2 Number of Switches between AOIs

A 3x3 mixed ANOVA (condition: no-conflict, conflict, neutral; group: no feedback, color feedback, double feedback) for the number of switches between the areas of interest (graphic, text) was conducted with item-wise arranged data. Condition was between-subject factor whereas group was within-subject factor. The results revealed a main effect of condition ($F_{(2, 15)} = 10.04$, $p < .001$, $\eta_p^2 = .57$). There were more switches between the AOIs (graph, text) for the no-conflict problems ($M = 5.93$, $SE = .43$) than conflict problems ($M = 3.89$, $SE = .43$) and neutral ($M = 3.27$, $SE = .43$).

4.2.3 Total Fixation Duration

A 3x2x3 mixed ANOVA (condition: no-conflict, conflict, neutral; area: graphic, text; group: no feedback, color feedback, double feedback) for the total fixation duration was conducted with item-wise arranged data. Group and AOI were within-subject factors whereas condition was a between-subject factor. The results revealed a main effect of area ($F_{(1, 15)} = 352.01$, $p < .001$, $\eta_p^2 = .95$) and condition ($F_{(2, 15)} = 9.55$, $p < .05$, $\eta_p^2 = .56$) an area*condition interaction ($F_{(2, 15)} = 4.43$, $p < .05$, $\eta_p^2 = .37$) and a group*area interaction ($F_{(2, 30)} = 12.45$, $p < .001$, $\eta_p^2 = .45$). The results of the test demonstrated that the total fixation duration for the texts ($M = 11.59$, $SE = .34$) was much longer than for the graphics ($M = 3.46$, $SE = .17$). Moreover, total fixation duration for the neutral problems ($M = 8.77$, $SE = .3$) was revealed to be longer than the no-conflict ($M = 6.9$, $SE = .3$) and conflict problems ($M = 6.86$, $SE = .3$). In the graphic area, total fixation duration was the highest for the neutral condition ($M = 4.11$, $SE = .31$), followed by the conflict condition ($M = 3.58$, $SE = .31$) whereas it was the lowest for the no-conflict condition ($M = 2.69$, $SE = .31$). In addition, for the text area; participants tended to fixate longer for the neutral problems ($M = 13.44$, $SE = .59$) than for the no-conflict problems ($M = 11.18$, $SE = .59$) and even shorter for the conflict condition ($M = 10.14$, $SE = .59$). The comparison of the total fixation durations of the three groups (No Feedback, Color Feedback, and Double Feedback) in terms of area showed that the means for the total fixation duration of the graphic area was the highest for the Color Feedback Group ($M = 3.64$, $SE = .26$), followed by the Double Feedback Group ($M = 3.55$, $SE = .25$) and lastly by the No Feedback Group ($M = 3.18$, $SE = .26$) which had the lowest values. Moreover, the groups differed in terms of total fixations times in the text area. Subjects in the No Feedback Group ($M = 12.21$, $SE = .5$) spent more time fixated on the text than the Color Feedback Group ($M = 11.81$, $SE = .45$) and lastly the Double Feedback Group ($M = 10.75$, $SE = .56$).

4.2.4 Total Visit Duration

A 3x2x3 mixed ANOVA (condition: no-conflict, conflict, neutral; area: graphic, text; group: no feedback, color feedback, double feedback) for the total visit duration was conducted with item-wise arranged data. Group and area were within-subject factors whereas condition was between-subject factor. The results revealed a main effect of area ($F_{(1, 15)} = 380.74$, $p < .001$, $\eta_p^2 = .96$), an area*condition interaction ($F_{(2, 15)} = 3.84$, $p < .05$, $\eta_p^2 = .33$) and a group*area interaction ($F_{(2, 30)} = 5.49$, $p < .01$, $\eta_p^2 = .26$). The results showed that the total time participants visited the text area ($M = 11.59$, $SE = .34$) was longer than the total visit duration of the graphic area ($M = 3.46$, $SE = .17$). For the graphic area, total visit duration was the highest for the neutral condition ($M = 4.11$, SE

=.31), somewhat lower for the conflict condition (*M* =3.58, *SE* =.31) came as second and lastly it was the lowest for the no- conflict condition (*M* =2.69, *SE* =.31). In addition, for the text area; participants tended to visit the neutral problems longer (*M* =13.44, *SE* =.59) than no-conflict problems (*M* =11.18, *SE* =.59). Moreover, they seemed to visit the text in the conflict condition (*M* =10.14, *SE* =.59) less than in the other conditions. For the graphic area, total visit duration was the highest for the Color Feedback Group (*M* =3.64, *SE* =.26), somewhat lower for the Double Feedback Group (*M* =3.55, *SE* =.25) and lowest for the No Feedback Group (*M* =3.18, *SE* =.26). In addition, for the text area; participants tended to have longer total visit durations in the No Feedback Group (*M* =12.21, *SE* =.5) than Color Feedback Group (*M* =11.81, *SE* =.45). Moreover, they seemed to visit the text in the Double Feedback Group (*M* =10.75, *SE* =.56) less than in the other groups.

4.2.5 Average Visit Duration

A 3x2x3 mixed ANOVA (condition: no-conflict, conflict, neutral; area: graphic, text; group: no feedback, color feedback, double feedback) for the average visit duration was conducted with item-wise arranged data. Group and area were within-subject factors whereas condition was between-subject factor. The results revealed a main effect of area *($F_{(1, 15)}$ = 102.94, $p < .001$, $\eta_p^2 = .87$)* and a group*area interaction *($F_{(2, 30)}$ = 4.65, $p < .05$, $\eta_p^2 = .23$).* The results showed that average visit duration for the text area (*M* =2.97, *SE* =.08) was longer than the average visit duration for the graphic area (*M* =1.09, *SE* =.16).

For the graphic area, average visit duration was the highest for the Double Feedback Group (*M* =1.42, *SE* =.46), somewhat lower for the Color Feedback Group (*M* =.99, *SE* =.06) and lowest for the No Feedback Group (*M* =.85, *SE* =.06). In addition, for the text area; participants tended to have longer average visit durations in the Color Feedback Group (*M* =3.24, *SE* =.12) than in the No Feedback Group (*M* =3.07, *SE* =.11). Moreover, they seemed to visit the text in the Double Feedback Group (*M* =2.6, *SE* =.09) less than in the other groups.

4.2.6 Visit Count

A 3x2x3 mixed ANOVA (condition: no-conflict, conflict, neutral; area: graphic, text; group: no feedback, color feedback, double feedback) for the visit count was conducted with item-wise arranged data. Group and AOI were within-subject factors whereas condition was a between-subject factor. The results revealed a main effect of area *($F_{(1, 15)}$ = 39.46, $p < .001$, $\eta_p^2 = .72$)* and a group*area interaction *($F_{(2, 30)}$ = 7.71, $p < .01$, $\eta_p^2 = .34$).* The results revealed that subjects visited the text area (*M* =2.91, *SE* =.07) more frequently than the graphic area (*M* =2.67, *SE* =.08).

For the graphic area, visit count was the highest for the Double Feedback Group (*M* =2.72, *SE* =.13), somewhat lower in the No Feedback Group (*M* =2.69, *SE* =.11) and lowest in the Color Feedback Group (*M* =2.61, *SE* =.12). In addition, for the text area; participants tended to have more visits in the Double Feedback Group (*M* =3.1, *SE* =.14) than in the No Feedback Group (*M* =2.98, *SE* =.12). Moreover, they seemed to visit the text in the Color Feedback Group (*M* =2.66, *SE* =.11) less than in the other groups.

CHAPTER 5

DISCUSSION

5.1 Discussion of the Results

The main goal of this study was to investigate to what extent providing feedback for questions about frequency distributions along with the direct experience of random sampling, have an effect on reasoning about base-rate problems. Participants in the No Feedback Group experienced the random sampling process (eighteen times; once before each item); they were presented with the color questions (eighteen times; once before each item); and on the final screen they saw the frequency distribution of the population in the problem on the one side of the screen (left or right) along with the base-rate problem on the other side of the screen (left or right) eighteen times. In the Color Feedback group the procedure was the same except the fact that the participants for this group were chosen according to their performances in a separate short pilot experiment which was conducted before the actual experiment. In this short pilot experiment, the participants were presented with eight questions of frequency distributions about color groups (it was the same color question used in the other experiments) after they had chosen one individual randomly out of ten faces (it was the same stimuli used in other experiments where there were ten neutral faces). In addition, they were provided with a feedback on the accuracy of their answers to the color questions. Participants who did at least six questions right were selected as they had learned how to answer the color question correctly and were included in the experiment for the Color Feedback Group. Finally, subjects in the Double Feedback group were presented with the same eighteen base-rate problems in the same procedure but this time, for the first nine questions (equal number of items from each condition were used) they were provided with feedback for both the color questions and the scenario questions.

5.1.1 E-Prime Results

Accuracy scores and reaction times of judgments obtained via E-Prime 2.0 were analyzed. In general, it was hypothesized that providing the participants with feedback would cause a decrease in the base-rate neglect. In this sense, the Double Feedback group was expected to be the most successful one in terms of the accuracy scores for scenario questions compared to the other two groups and the Color Feedback group was hypothesized to have higher accuracy scores than No Feedback group in terms of both the color questions and the scenario questions. In addition, problems in the no-conflict condition were expected to be solved with higher accuracy scores than neutral and conflict problems.

The results of the analyses revealed that the No Feedback Group had the lowest accuracy scores for scenario questions and the participants in the Color Feedback Group – although more successful than the former – were not as successful as the participants in the Double Feedback Group. This difference can be explained with the provided feedbacks on the color questions and the scenario questions in the Double Feedback Group. In addition the reason for the Color Feedback Group's higher accuracy scores might be due to the short training session that they had done before the experiment. With the help of the feedbacks that they were presented after the eight color questions, they might have learnt how to do the color questions and then applied this knowledge to the scenario questions. The results in terms of condition are in line with the literature; participants were less accurate in their judgments for the conflict problems than they were for the neutral problems and no-conflict problems overall.

The results of the accuracy scores of the color questions demonstrated that participants in the Color Feedback Group were more successful than participants in the No Feedback Group and Double Feedback Group. It is easily explainable why the Color Feedback Group was the best since the participants in this group did a short training session and learned how to do the color questions. On the other hand, it is hard to explain why the Double Feedback Group was less successful than the No Feedback Group. One explanation might be that the participants for the Color Feedback Group were chosen as a result of the training session but there was no selection process for the participants in the Double Feedback Group. Moreover, getting feedbacks for both the color questions and the scenario questions might have caused confusion because of neutral and conflicting items since both conditions required deeper investigation of the frequency distributions.

In a separate analysis within the Double Feedback Group it was revealed that participants were more successful in their answers for the no-conflict problems and neutral problems than for the conflict problems as it was in the general analysis. Moreover, its training and test parts were compared in terms of the accuracy scores of the scenario questions and more accurate results were found in the test part than in the trial part. This difference can be explained with the feedbacks that were given in the first half of the experiment.

The reaction time analysis revealed that participants made judgments for the items in the neutral condition slower than they did in the no-conflict condition and conflict condition. In the present study the participants might not have detected a conflict between the base-rates displayed via the graphic and the diagnostic information or they might have detected the conflict but chose not to study the graphic and decided to rely on the verbal descriptions. On the other hand, it is reasonable for them to spend more time on the neutral items in which there was no clue about the category to which the person was most likely to belong. This finding is different from the preceding studies (De Neys & Glumicic, 2008) where the response times for the conflict condition and neutral condition were found to be longer than for the no-conflict condition.

5.1.2 Eye-Tracking Results

Eye-tracking data analyses consisting of various dependent measures: first looks/lasts looks in the specified areas of interest, number of switches between the two areas of interest, total fixation duration, total visit duration and average visit duration were investigated. It was hypothesized that the number of switches between the AOIs (graphic, text) would be lower for the base-rate problems in the no-conflict condition than neutral condition and conflict condition would have the most number of switches. Moreover, in the graphic area, total fixation duration, total visit duration and visit count were expected to be longer for the base-rate problems in the conflict condition than in the text area. In the current analyses, although total fixation duration and total visit duration revealed somewhat similar results, none of the measures in this study was redundant.

The sign test revealed that participants looked at the text area first more frequently than they looked at the graphic area in general. Moreover, there were more last looks in the text area than there were in the graphic area in general. These findings, however, may not indicate participants' neglect of attending the graphics where the crucial frequency information was displayed.

In terms of the number of switches between the graphic and the text areas, no-conflict problems had the highest scores. As for the other two conditions, there were more switches in the conflict problems than in the neutral problems. For the no-conflict problems the result might have occurred due to the fact that the participants might have noticed if there were graphics presented with the problem, they might have something to do with the solution. This may be the reason why they went back and forth between the AOIs more often, namely with the aim of figuring out the role of the graphic in the solution. On the other hand, in the conflict condition the subjects may have seen the graphics as redundant.

The analyses of the total fixation duration, total visit duration, average visit duration and visit count demonstrated a main effect of area; all the three duration metrics and visit count revealed higher rates for the text area rather than the graphic area.

In terms of total fixation duration the participants spent more time on the neutral problems than no-conflict problems and conflict problems. For the participants, neutral problems required more fixation time in order to investigate the relation since there was no cue in the description as to which group the protagonist might belong. The reason why the fixation duration was less for conflict problems might be participants' lack of experience of conflict. In other words, even though they detected the conflict, they might have chosen to ignore it. Either they did not experience any conflict and assigned the protagonist to the more similar group straightforwardly, or they detected the conflict but ignored it rather than trying to resolve it with conscious effort.

In the graphic area total fixation duration and total visit duration were the longest for the neutral condition followed by the conflict condition whereas they were the lowest for the no-conflict condition. This result can be explained along the same lines with the previous

explanation. The subjects studied the graphic part for the neutral problems and the conflict problems longer because they sought the crucial information there but spent less time on fixating and visiting it in the no-conflict condition because there the graphics information was redundant. For the text area both metrics revealed the same results again; the participants tended to fixate longer for the neutral problems than for the no-conflict problems and even shorter for the conflict condition. At least the neutral condition had more time spent on the graphic as well as on the text as compared to the other two conditions.

With respect to experimental groups, total fixation duration and total visit duration of the graphic area was the highest for the Color Feedback Group, followed by the Double Feedback Group and lastly by the No Feedback Group which had the lowest values. The participants might have understood that the graphics were important for the solution in the Color Feedback Group because of the training that they had undergone before the main experiment. It also makes sense for the Double Feedback Group since they got feedbacks on both types of questions in the first half of the experiment. Participants in the No-feedback Group, however, might have skipped considering the graphic information because they were not aware that some crucial information was hidden there.

In terms of total fixation duration and total visit duration, the results showed that the No Feedback Group spent more time fixating on the text than the Color Feedback Group and lastly the Double Feedback Group. This result can be explained in terms of the importance that subjects gave to the text in the various groups: there was no clue for the role of the graphic part, as in the No-Feedback group, they would rely more on the text and consider the graphic information as redundant or irrelevant.

On the other hand the results for the average visit duration and visit count were somewhat different from the other measurements. The Group*area interaction for the average visit duration in the graphic area revealed the highest average duration for the Double Feedback Group, somewhat lower durations for the Color Feedback Group and the lowest ones for the No Feedback Group. This result in general is the most compatible one with the accuracy scores in which the same order of groups was revealed: the Double Feedback Group was the most successful one and the scores for the No Feedback Group was somewhat lower than for the Color Feedback Group. It can be inferred that in the groups where feedbacks was provided at least for one type of question (color question) or both questions (color and feedback question) subjects tended to visit the graphic area longer on average than subjects in the group where there was no feedback. The analysis for the visit count produced the result that the Double Feedback Group had the highest number of visits in the graphic area.

Finally, the results for the average visit duration in the text area showed that the participants spent least time on the text in the Double Feedback Group. This mirrors the fact that they spent more time on the graphics area, i.e., they allocated more time relative to the graphical than to the text areas as compared to the other groups. The results for the visit count on the other hand demonstrated that the participants tended to have more visits in the text area in the Double Feedback Group than in the No Feedback Group.

Moreover, they seemed to visit the text in the Color Feedback Group less than in the other groups. It seems like in order to make a more accurate judgment the subjects relied on text along with the graphic in the Double Feedback Group.

In general, even though some results from the eye-tracking data were hard to interpret and need further and deeper investigation; our manipulation of the feedback across the different experimental groups seems to have a facilitating effect on the resolution of base-rate problems. Providing a short training part for the Color Group helped the participants to understand the role of frequency distributions for the problems better. Since the successful participants from the training session were chosen for the Color Feedback Group, they might have transferred the relation that they have learned across the trials with color questions and feedbacks to the judgments of the scenario questions. This explains why this group had higher accuracy scores for the scenario questions than the No Feedback Group. Moreover, giving the participants feedback after both the color questions and scenario questions in the first part of the experiment in the Double Feedback Group may be the reason why this group succeeded the most. This generalization can be supported with the eye-tracking results in which average visit durations for the graphic part in the Double Feedback Group was longer than in the Color Feedback Group and the No Feedback Group. For further studies these base-rate problems can be examined in detail by determining more than two areas of interest; one for the graphic, one for the diagnostic information part and the last one for the category choices (options for the answer) (Ball *et al.*, 2006; De Neys & Glumicic, 2008).

Eye-tracking is a frequently used methodology in the studies of human cognitive processes. In the light of the present study, it proved to be a useful methodology to study online decision making processes. For example; the focus of attention can be investigated by this method since it is associated with the fixation area and it provides an insight into how the data is being processed at a given time. In the literature the studies that measured fixation duration revealed that processing levels increases as fixation duration increases (Pomplun, Ritter, & Velichkovsky, 1996; Rayner, 1998; Velichkovsky, 1999; Velichkovsky, Rothert, Kopf, Dornhofer, & Joos, 2002). This may be parallel with this study's finding that, in neutral problems – because there was ambiguity – the subjects needed to spend more time on them. In this respect, a major contribution of this thesis to the literature may be the finding that participants reasoned the better the more they looked to the graphic area where the crucial information was provided, especially in the conflict condition. Mainly base-rate neglect is revealed in looking to the text at the expense at looking at the graphics like it was the case for the No Feedback Condition. Even though it is hard to set a direct relationship between solving the problems right, experiencing the direct and repetitive sampling from the same population, receiving feedback and understanding the role of frequency distributions, it can be concluded that these manipulations caused only a modest difference between the three experimental groups (No Feedback, Color Feedback and Double Feedback).

5.2 Limitations and Future Research

5.2.1 Limitations of the Study

First of all even though the participants from METU were majoring in varying subjects; the students from certain departments such as, statistics and mathematics should have been separated or we might have added another factor into the analyses namely "education".

Secondly, the subjects might have been tested with appropriate material in order to control the internal factors such as working memory capacity and intelligence. Moreover, they might have been asked about their genetic history if it was possible.

Third, the participants in the Double Feedback Group were not chosen in terms of their accuracy scores in the first phase of the experiment while the subjects in the Color Feedback Group were selected according to their success levels in terms of the accuracy scores in the training part of the experiment.

Finally, the number of items in each condition for the training phase and the test phase may not be enough. In other words; the subjects might have needed more than nine items (three items from each condition) in the training phase in order to learn how to solve the base-rate neglect problems.

5.2.2 Suggestions for Further Research

For future direction, a questionnaire may be given to the participants right after the experiments consisting of questions examining the level of introspection about the base-rate neglect problems.

Secondly, more than two areas of interest may be choosen for a more detailed investigation for the eye-tracking study. There should be at least three parts in the screen; the first part where there is diagnostic information, the second part with the question and answer categories and finally the third part which is composed of the frequency distribution and the names of the categories.

Thirdly, the probability values used in the pilot study which had low uncertainity and the probability values in the main study that had higher uncertainity should be compared in a future research to see how they affect our reasoning about base-rate neglect problems.

REFERENCES

Baddeley, A. D. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology*, *63*, 537–546.

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, *53*, 77–86.

Barbey, A. K. & Sloman, S. A. (2007). Base-rate respect: From statistical formats to cognitive structures. *Behavioral and Brain Sciences, 30,* 287-298.

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), *Decision making under uncertainty (pp.* 39-61). Amsterdam: North-Holland.

Bar-Hillel, M., & Fischhoff, B. (1981). When do base rates affect predictions? *Journal of Personality and Social Psychology, 41,* 671-680.

Birnbaum, M, H. and Mellers, B. A. (1983). Bayesian inference: combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45,* 792-804.

Bousfield, W. A., & Sedgewick, C. H. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, *30,* 149-165.

Chapman, Loren J. and Jean P. (1969). Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs. *Journal of Abnormal Psychology*, *74* (3): 271–80.

Cohen, J. (1962). The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences, 4*, 317-331.

Cooper, W. S. (1989). How evolutionary biology challenges the classical theory of rational choice. *Biology and Philosophy*, *4*, 457-481.

Cummins, D.D. (2004) The evolution of reasoning. In J.P. Leighton and R.J. Sternberg (eds.), *The nature of reasoning* (pp. 339-374). Cambridge: Cambridge University Press.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113*, 45-61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of reasoning. *Cognition, 106,* 1248-1299.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory. Processes of judgment and choice. *Annual Review of Psychology, 32*, 53-88.

Epley, N. (2004). A tale of Tuned Decks? Anchoring as accessibility and anchoring as adjustment. In D. J. Koehler, & N. Harvey (Eds.), The Blackwell Handbook of Judgment and Decision Making (pp. 240-256). Oxford, UK: Blackwell Publishers.

Epley, N., & Gilovich, T. (2005). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*, 311-318.

Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451-468.

Evans, J. St. B. T. & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223-241, 263-271.

Evans, J. St. B. T., & Over, D. E. (1996). Rationality and reasoning. Hove, England: Psychology Press.

Evans, J. St. B. T., Handley, S. J., Over, D. E. & Perham, N. (2002) Background beliefs in Bayesian inference. *Memory & Cognition, 30*:179–90.

Fischhoff B. (1982). Debiasing. In: Kahneman D, Slovic P, Tversky A, eds. Judgment under Uncertainty: Heuristics and Biases (pp. 422-444) Cambridge, MA: Cambridge University Press.

Fischhoff. B.. Slovic, P. and Lichtenstein. S. (1979). Improving intuitive judgment by subjective sensitivity analysis. *Organizational Behavior and Hutnan Performance. 231,* 339-359.

FischofT, B., Slovic, P. & Lichtenstein, S. (1979) Subjective sensitivity analysis. *Organizational Behavior and Human Performance* 23:339-59.

Fong GT, Krantz DH, Nisbett RE. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*. *18*:253-292.

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning. 15*:2, 105-128.

Gigerenzer, G, and Hoffrage, U. (1995). How to Improve Bayesian reasoning without instruction: frequency formats. *Psychological Review. 10*, 684-704.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology (Vol. 2,* pp. 83-115). Chichester, England: Wiley.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology and vice versa. In G. Wright & P. Avion (Eds.), *Subjective probability* (pp. 129-162). New York: Wiley.

Gigerenzer, G., & Murray, D. J. (1987). Cognition as intuitive statistics. Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). The empire of chance: How probability changed science and everyday life. Cambridge: Cambridge University Press.

Gigerenzer, G.. Hell, W. and Blank, H. Presentation and content: the use of base rates as a continuous variable. (1988). *Journal of Experimental Psychology: Human Perception and Performance. 14,* 513-525.

Ginossar, Z. and Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology. 16,* 228-242.

Goodie, S. A. and Crooks, L. C. (2004). Time-Pressure Effects on Performance in a Base-Rate Task. *The Journal of General Psychology. 131*:1, 18-28.

Griffin, D., Buehler, R. (1999). Frequency, Probability, and Prediction: Easy Solutions to Cognitive Illusions? *Cognitive Psychology, 38,* 48-78.

Hacking, I. (1975). The emergence of probability. Cambridge: Cambridge University Press.

Hockey, G. R. J. (1993). Cognitive-energetical control mechanisms in the management of work demands and psychological health. In A. D. Baddeley & L. Weiskrantz (Eds.), Attention: Selection, awareness, and control: A tribute to Donald Broadbent. Oxford: Clarendon Press/Oxford University Press.

Janis, I. L., & Mann, L. (1977). Emergency decision making: A theoretical analysis of responses to disaster warnings. *Journal of Human Stress*, *3*, 35–48.

Kahneman, D., & Tversky, A. (2000). Conflict resolution: A cognitive perspective. In D. Kahneman and A. Tversky (Eds.), Choices, values, and frames (pp. 473–487). Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgement of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review,* Vol 102(3), 490-502.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral & Brain Sciences*, *19*, 1-53.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1133-1145.

Kosonen, P., & Winne, P.H. (1995). Effects of teaching statistical laws on reasoning about everyday problems. *Journal of Educational Psychology*, *87*, 33–46.

Krosnick, J. A., Li, E, & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology, 59,* 1140-1152.

Kunda, Ziva (1999). *Social Cognition: Making Sense of People*. MIT Press.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology, 1, 65-82.*

Manis, M., Dovalina, I., Avis, N. E. & Cardoze, S. (1980) Base rates can affect individual predictions. *Journal of Personality and Social Psychology, 38, 231- 48.*

MB Welsh and DJ Navarro (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes, 119*, 1-14.

McGuire, T. W. & Melone, N. P. (1992) Conservatism in integrating information: Implications for international management. In: *Business behavior and information*, ed. Y. Ijiri & I. Nakano. Carnegie Mellon University Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Fong, G. T. (1982). Improving inductive inference. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Osman M (2004) . An evaluation of dual-process theories of reasoning**.** *Psychon Bull Rev* vol. *11*(6), 988-1010.

Over, D., & Evans, J. S. B. T. (1997). Two cheers for deductive competence. *Current Psychology of Cognition*, *16*, 255-278.

Parmley M. C. (2006), 'Effects of the Confirmation Bias on Diagnostic Decision Making', Unpublished doctoral dissertation, Drexel University.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (in press). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition, 123,* 335–346.

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception, 25*, 931–948.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition, 1*, 56-60.

Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review, 104*, 603-617.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Schwarz, N., Strack, E, Hilton, D. J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition, 9*, 67-83.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379-398). Cambridge: Cambridge University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral & Brain Sciences*. *23*, 645-665.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421-440). Cambridge: Cambridge University Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695.

Stephens, D. W. & Krebs, J. R. (1986) *Foraging theory*. Princeton University Press. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*, 145–156.

Tversky, A. & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology, 5* (2), 677-695.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), Progress in social psychology (pp. 49-72). Hillsdale, NJ: Erlbaum.

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

Tversky, A., & Kahneman, D. (1971). The belief in the law of small numbers. *Psychological Bulletin*. *76*, 105-110.

Velichkovsky, B. M. (1999). From levels of processing to stratification of cognition: Converging evidence from three domains of research. In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 203–235). Amsterdam, Netherlands: John Benjamins Publishing Company.

Velichkovsky, B.M., Rothert, A., Kopf, M., Dornhoefer, S.M. & Joos, M. (2002). Towards an express diagnostics for level of processing and hazard perception. *Transportation Research, Part F. 5*(2), 145-156.

Payne, J.W. (1980). Information processing theory: Some concepts and methods applied to decision research. In T. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, N.J.: Lawrence Erlbaum.

Wells. G. L. and Harvey, J. H. (1977). “Do people use consensus information in making causal attributions?” *Journal of Personality and Social Psychology, 35*, 279-293.

APPENDICES

APPENDIX A: BASE-RATE PROBLEM POOL

No-conflict Problems

1. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 996 kişinin National Geographic kanalını, 4'ünün ise Kanal D'yi seyretmeyi sevdiği ortaya çıktı. Ayşegül bu çalışmadan ratgele seçilen bir katılımcıdır.

Ayşegül 35 yaşındadır. O, bir dergi için eleştiri yazıları yazar. Kocası bir üniversitede çalışır. Ayşegül resim yapmayı ve fotoğraf çekmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Ayşegül National Geographic kanalını seyretmeyi sever.

2- Ayşegül Kanal D'yi seyretmeyi sever.

2. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 996 kişinin köyde, 4'ünün ise şehirde yaşadığı ortaya çıktı. Müge bu çalışmadan rastgele seçilen bir katılımcıdır.

Müge 22 yaşındadır. Ata biner. Okuldan sonra evde hayvanlara bakar. Hafta sonları erken kalkıp büyük anne ve büyük babasını ziyaret eder.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Müge köyde yaşar.

2- Müge şehirde yaşar.

3. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5 kişi kadın iken, 995 kişi erkekti. Deniz bu çalışmadan rastgele seçilen bir katılımcıdır.

Deniz 32 yaşında, kendine güvenen hırslı bir kimsedir. Amacı kariyer yapmaktır. Aynı zamanda çok spor yapar ve oldukça kaslı bir vücuda sahiptir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Deniz erkektir.

2- Deniz kadındır.

4. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 995 kişi on altı yaşında iken, 5 kişi kırk yaşındaydı. Armağan bu çalışmadan rastgele seçilen bir katılımcıdır.

Armağan tekno ve elektronik müzik dinlemeyi sever. Genellikle dar kazak ve kot pantolon giyer. Dans etmeyi sever ve küçük bir burun piercingi var.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Armağan on altı yaşındadır.

2- Armağan kırk yaşındadır.

5. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 995 kişi kıyafetlerini yüksek kalitedeki mağazalardan alırken, 5 kişi semt pazarından alıyordu. Yağmur bu çalışmadan rastgele seçilen bir katılımcıdır.

Yağmur 33 yaşında, bir muhasebe ofisinde çalışan ve Porsche kullanan bir kadndır. O, erkek arkadaşı ile şık bir çatı katında yaşamaktadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Yağmur kıyafetlerini yüksek kalitedeki mağazadan alır.

2- Yağmur kıyafetlerini semt pazarından alır.

6. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997 kişi kadın iken 3 kişi erkekti. Erin bu çalışmadan rastgele seçilen bir katılımcıdır.

Erin 13 yaşındadır ve en çok sanatla ilgilenir. O en çok alışveriş yapmayı ve arkadaşlarında yatıya kalıp okuldaki diğer çocuklarla ilgili dedikodu yapmayı sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Erin kadındır.

2- Erin erkektir.

7. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'sinde dövme varken, 3 kişide yoktur. Poyraz bu çalışmadan rastgele seçilen bir katılımcıdır.

Poyraz 29 yaşında kısa bir süre hapishanede yatmış bir erkektir. Son 2 yıldır tek başına yaşamaktadır. Eski bir arabaya sahiptir ve punk müzik dinlemektedir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Poyrazın dövmesi vardır.

2- Poyrazın dövmesi yoktur.

8. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 996'sı anaokulu öğretmeni iken, 4'ü idari yöneticidir. Selin bu çalışmadan rastgele seçilen bir katılımcıdır.

Selin 37 yaşındadır. O, evlidir ve 3 çocuğa sahiptir. Kocası veterinerdir. Kendini ailesine adamıştır ve her gün çocuklarıyla birlikte çizgi film izler.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Selin anaokulu öğretmenidir.

2- Selin idari yöneticidir.

9. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü Rolling Stones hayranı iken 996'sı Britney Spears hayranı idi. Ceren bu çalışmadan rastgele seçilen bir katılımcıdır.

Ceren 15 yaşındadır. O, arkadaşlarıyla alışveriş merkezine gitmeyi ve arkadaşlarıyla okuldaki hoşlandıkları kişiler hakkında konuşmayı sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Ceren Rolling Stones hayranıdır.

2- Ceren Britney Spears hayranıdır.

10. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5 kişi Amerikalı iken 995 kişi Fransızdır. Martine bu çalışmadan rastgele seçilen bir katılımcıdır.

Martine 26 yaşındadır. O, iki dil bilir ve boş zamanlarında çok okuma yapar. Çok şık giyinir ve harika bir aşçıdır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Martine Amerikalıdır.

2- Martine Fransızdır.

11. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5 kişi İsveçli iken 995 kişi İtalyandır. Marco bu çalışmadan rastgele seçilen bir katılımcıdır.

Marco 16 yaşındadır. Arkadaşlarıyla futbol oynamayı ve sonrasında hep beraber pizza yemeye gitmeyi ya da birinin evinde ev yapımı makarna yemek için toplanmayı çok sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Marco İsveçlidir.

2- Marco İtalyandır.

12. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 3'ü kırk yaşında iken, 997'si 17 yaşındaydı.

Umur İzmir'de yaşamaktadır. Her gün yakın arkadaşlarıyla vakit geçirir ve MTV izlemeyi sever. O bir Metallica hayranıdır ve kendi arabasını alabilmek için para biriktirmektedir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Umur kırk yaşındadır.

2- Umur on yedi yaşındadır.

Conflict Problems

13. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4 kişinin ikinci el Renault, 996 kişinin ise BMW kullandığı ortaya çıktı. Arda bu çalışmadan rastgele seçilen bir katılımcıdır.

Arda 38 yaşındadır. O, bir çelik fabrikasında çalışır. Ankara'nın kenar mahallelerinden birinde ufak bir dairede yaşar. Karısı onu terketmiştir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Arda BMW kullanır.

2- Arda Renault kullanır.

14. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 995'i Müslüman iken, 5 kişi Budist idi. Gizem bu çalışmadan rastgele seçilen bir katılımcıdır.

Gizem 19 yaşındadır. Felsefe ile uğraşmayı sever ve materyalizmden hiç hoşlanmaz. İkinci el kıyafetler giyer ve bir gün Hindistan'a gitmeyi çok istemektedir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Gizem Budisttir.

2- Gizem Müslümandır.

15. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i on beş yaşında iken, 995'i elli yaşındaydı. Gökçe bu çalışmadan rastgele seçilen bir katılımcıdır.

Gökçe haftasonları arkadaşları ile dansa gitmeyi sever. Genellikle hazır yiyecekleri tercih eder ve göbeğinde küçük bir piercing vardır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Gökçe on be yaşındadır.

2- Gökçe elli yaşındadır.

16. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 994'ü İsveçli iken, 6'sı İtalyandı. Mario bu çalışmadan rastgele seçilen bir katılımcıdır.

Mario yirmi beş yaşındadır. Genç, yakışıklı ve aynı zamanda çapkın bir adamdır. En sevdiği yemek annesinin yaptığı makarnadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Mario İsveçlidir.

2- Mario İtalyandır.

17. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i mühendis iken 995'i avukat idi. Berk bu çalışmadan rastgele seçilen bir katılımcıdır.

Berk 36 yaşındadır. O, bekardır ve bir miktar içine kapanıktır. Boş zamanlarını bilim-kurgu okuyarak ve bilgisayar programları yazarak geçirmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Berk mühendistir.

2- Berk avukattır.

18. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü erkek iken 996'sı kadındır. Evrim bu çalışmadan rastgele seçilen bir katılımcıdır.

Evrim 23 yaşındadır ve mühendislik fakültesinden mezun olmak üzeredir. Cuma akşamları, Evrim arkadaşlarıyla dışarı çıkmayı, onlarla birlikte gürültülü müzik eşliğinde bira içmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Evrim erkektir.

2- Evrim kadındır.

19. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 3'ü apartman dairesinde yaşarken, 997'si çiftlik evinde yaşıyordu. Ali bu çalışmadan rastgele seçilen bir katılımcıdır.

Ali büyük ve başarılı bir şirkette çalışır ve bekardır. Uzun saatler boyunca çalışır ve işe giderken Armani takım elbiselerinden giyer. Güneş gözlüğü kullanmayı sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Ali bir apartman dairesinde yaşamaktadır.

2- Ali bir çiftlik evinde yaşamaktadır.

20. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si hemşire iken 3'ü doktordu. Yağız bu çalışmadan rastgele seçilen bir katılımcıdır.

Yağız 34 yaşındadır. Lüks bir banliyöde güzel bir evde yaşamaktadır. O, hoşsohbettir ve politikayla yakından ilgilidir. Vaktinin çoğunu kariyerine yatırım yaparak geçirir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Yağız hemşiredir.

2- Yağız doktordur.

21. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ünün favori televizyon dizisinin Star Trek, 996'sının ise Dallas olduğu ortaya çıktı. Özgür bu çalışmadan rastgele seçilen bir katılımcıdır.

Özgür 26 yaşında, fizik alanında lisansüstü eğitimine devam eden bir adamdır. O, vaktinin çoğunu evde oturup video oyunları oynayarak geçirmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Özgür'ün favori televizyon dizisi Star Trek'tir.

2- Özgür'ün favori televizyon dizisi Dallas'tır.

22. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i on altı yaşında iken, 995'i elli yaşında idi. Elif bu çalışmadan rastgele seçilen bir katılımcıdır.

Elif hip hop ve rap tarzı müzik dinlemeyi sever. Genellikle kısa etekler giyer ve arkadaşlarıyla sabaha kadar dans edebileceği partilere katılır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Elif on altı yaşındadır.

2- Elif elli yaşındadır.

23. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i mühendis iken, 995'i avukattı. Göker bu çalışmadan rastgele seçilen bir katılımcıdır.

Göker 45 yaşındadır ve dört çocuğu vardır. Genellikle muhafazakardır, politik ve sosyal konularla hiç ilgilenmez. Yelkenle denize açılmayı ve matematiksel bilmeceleri sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Göker mühendistir.

2- Göker avukattır.

24. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i psikolog iken, 995'i işletmeci idi. Burçin bu çalışmadan rastgele seçilen bir katılımcıdır.

Burçin haftasonları arkadaşlarıyla gürültülü olmayan restoranlarda buluşup yemek

yemeyi çok sever. Hayvan barınaklarında gönüllü çalışır ve insanlarla olan ilişkilerine önem verir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Burçin psikologdur.

2- Burçin işletmecidir.

Neutral Problems

25. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 995'i elli yaşında iken, 5 kişi altmış yaşındadır. Umur bu çalışmadan seçilen rastgele bir katılımcıdır.

Umur farklı kültürlere çok meraklı bir kimsedir. Diğer ülkelerin yemeklerini denemeyi sever. Macaristan'daki tatilinden yeni döndü.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Umur altmış yaşındadır.

2- Umur elli yaşındadır.

26. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü saksafon, 996'sı davul çalmaktadır. Toprak bu çalışmadan seçilen rastgele bir katılımcıdır.

Toprak 19 yaşındadır. İstanbul'da okumaktadır ve kız arkadaşı yoktur. Biriktirdiği para ile eski, ikinci el bir araba satın almıştır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Toprak saksafon çalmaktadır.

2- Toprak davul çalmaktadır.

27. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i Arsenal, 995'i ise Real Madrid taraftarıdır. Emir bu çalışmadan rastgele seçilen bir katılımcıdır.

Emir 39 yaşındadır. Koyu bir futbol takipçisidir. Takımı kaybettiğinde haftaya kötü başlar. Oğlunu takımın kendi sahasında yaptığı her maça götürür.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Emir Arsenal taraftarıdır.

2- Emir Real Madrid taraftarıdır.

28. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 994'ü İzmirli, 6'sı ise Ankaralıdır. Karya bu çalışmadan rastgele seçilen bir katılımcıdır.

Karya on altı yaşındadır ve hala okula devam etmektedir. Seksen kilodur ve kendinden küçük dört yaşında bir kız kardeşi ile iki yıldır üniversitede okuyan kendinden büyük bir abisi vardır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Karya İzmirlidir.

2- Karya Ankaralıdır.

29. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i İstanbul Teknik Üniversitesi mezunu iken, 995'i Boğaziçi Üniversitesi mezunu idi. Tolga bu çalışmadan rastgele seçilen bir katılımcıdır.

Tolga 1.73 boyunda, esmer, iki küçük kız çocuğuna sahip bir babadır. O, üzeri tamamen posterlerle kaplı sarı bir karavan kullanmaktadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Tolga İstanbul Teknik Üniversitesi mezunudur.

2- Tolga Boğaziçi Üniversitesi mezunudur.

30. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 996'sı erkek iken, 4'ü kadındır. Bilge bu çalışmadan rastgele seçilen bir katılımcıdır.

Bilge 36 yaşında bir yazardır. İki erkek, bir de kız kardeşi vardır. O, koşmayı ve iyi filmler izlemeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Bilge erkektir.

2- Bilge kadındır.

31. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si biyoloji okurken, 3 kişi de kimya bölümünde okumaktadır. Yiğit bu çalışmadan rastgele seçilen bir katılımcıdır.

Yiğit 20 yaşındadır. İstanbul'da okumaktadır ve kalıcı bir kız arkadaşı yoktur. Biriktirdiği parayla henüz kendine ikinci el bir gitar alabilmiştir.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Yiğit biyoloji bölümünde okumaktadır.

2- Yiğit kimya bölümünde okumaktadır.

32. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si bilardo oyuncusu iken, 3 kişi de basketbol oyuncusuydu. Can bu çalışmadan rastgele seçilen bir katılımcıdır.

Can 29 yaşındadır ve hayatı boyunca New York'ta yaşamıştır. O, siyah saçlara ve yeşil gözlere sahiptir. Açık gri renkte bir araba kullanır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Can bilardo oyuncusudur.

2- Can basketbol oyuncusudur.

33. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü İzmir'de, 996'sı ise Ankara'da yaşamaktadır. Ayhan bu çalışmadan rastgele seçilen bir katılımcıdır.

Ayhan 28 yaşındadır. O, bir arkadaşıyla apartman dairesini paylaşmaktadır ve bir kız arkadaşı vardır. Basketbol izlemeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Ayhan İzmir'de yaşamaktadır.

2- Ayhan Ankara'da yaşamaktadır.

34. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i Bilgisayar Bilimleri'nden mezun iken, 995'i İngilizce'den mezu idi. Gülnur bu çalışmadan rastgele seçilen bir katılımcıdır.

Gülnur 20 yaşındadır ve Antalya'da şehir merkezinde yaşamaktadır. Onun en sevdiği yemek kıymalı makarnadır. Anne ve babası Muğla'da yaşamaktadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Gülnur Bilgisayar Bilimleri'nden mezun olmuştur.

2- Gülnur İngilizce'den mezun olmuştur.

35. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 3'ü yüksek lisansına Yale Üniversitesi'nde, 997'si ise Princeton'da devam etmekteydi. John bu çalışmadan rastgele seçilen bir katılımcıdır.

John 22 yaşındadır. Anne ve babası küçük kız kardeşi ile birlikte Toronto'da yaşamaktadır. John zamanının çoğunu kütüphanedeki klasik eserleri okuyarak geçirir. Bisiklet sürmeyi çok sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- John yüksek lisansına Yale Üniversitesi'nde devam etmektedir.

2- John yüksek lisansına Princeton Üniversitesi'nde devam etmektedir.

36. Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i Google Chrome'u kullanırken, 995'i İnternet Explorer'ı kullanıyordu. Özge bu çalışmadan rastgele seçilen bir katılımcıdır.

Özge 25 yaşındadır. Lisansını iktisat üzerine yapmıştır. Boş zamanlarında kitap okumayı, bilgisayar programları yazmayı ve internette oyun oynamayı sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1- Özge Google Chrome'u kullanmaktadır.

2- Özge Internet Explorer'ı kullanmaktadır.

APPENDIX B: THE STIMULI WITH PICTURES USED IN THE PILOT STUDIES

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 996 kişinin National Geographic kanalını, 4'ünün ise Kanal D'yi seyretmeyi sevdiği ortaya çıktı. Ayşegül bu çalışmadan rastgele seçilen bir katılımcıdır.

Ayşegül 35 yaşındadır. O, bir dergi için eleştiri yazıları yazar. Kocası bir üniversitede çalışır. Ayşegül resim yapmayı ve fotoğraf çekmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Ayşegül National Geografic kanalını seyretmeyi sever.
2. Ayşegül Kanal D'yi seyretmeyi sever.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5 kişi kadın iken, 995 kişi erkekti. Deniz bu çalışmadan rastgele seçilen bir katılımcıdır.

Deniz 32 yaşında, kendine güvenen hırslı bir kimsedir. Amacı kariyer yapmaktır.Aynı zamanda çok spor yapar ve oldukça kaslı bir vücuda sahiptir.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Deniz erkektir.
2. Deniz kadındır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 995 kişi kıyafetlerini yüksek kalitedeki mağazalardan alırken, 5 kişi semt pazarından alıyordu. Yağmur bu çalışmadan rastgele seçilen bir katılımcıdır.

Yağmur 33 yaşında, bir muhasebe ofisinde çalışan ve Porsche kullanan bir kadındır. O, erkek arkadaşı ile şık bir çatı katında yaşamaktadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Yağmur kıyafetlerini yüksek kalitedeki mağazadan alır.
2. Yağmur kıyafetlerini semt pazarından alır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997 kişi kadın iken 3'ü erkekti. Erin bu çalışmadan rastgele seçilen bir katılımcıdır.

Erin 13 yaşındadır ve en çok sanatla ilgilenir. O en çok alışveriş yapmayı ve arkadaşlarında yatıya kalıp okuldaki diğer çocuklarla ilgili dedikodu yapmayı sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Erin kadındır.
2. Erin erkektir.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcıların 997'sinde dövme varken, 3 kişide yoktur. Poyraz bu araştırmadan rastgele seçilen bir katılımcıdır.

Poyraz 29 yaşında, kısa bir süre hapishanede yatmış bir erkektir. Son 2 yıldır tek başına yaşamaktadır. Eski bir arabaya sahiptir ve punk müzik dinlemektedir.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Poyraz'ın dövmesi vardır.

2. Poyraz'ın dövmesi yoktur.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5 kişi İsveçli iken 995'i İtalyan'dır. Marco bu çalışmadan rastgele seçilen bir katılımcıdır.

Marco 16 yaşındadır. Arkadaşlarıyla futbol oynamayı ve sonrasında hep beraber pizza yemeğe gitmeyi ya da birinin evinde ev yapımı makarna yemek için toplanmayı çok sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Marco İsveçli'dir.

2. Marco İtalyan'dır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4 kişinin ikinci el Renault, 996 kişinin ise BMW kullandığı ortaya çıktı. Arda bu çalışmadan rastgele seçilen bir katılımcıdır.

Arda 38 yaşındadır. O, bir çelik fabrikasında çalışır. Ankara'nın kenar mahallelerinden birinde ufak bir dairede yaşar. Karısı onu terketmiştir.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Arda BMW kullanır.
2. Arda Renault kullanır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i on beş yaşında iken, 995'i elli yaşındaydı. Gökçe bu çalışmadan rastgele seçilen bir katılımcıdır.

Gökçe haftasonlarını arkadaşları ile dansa gitmeyi sever. Genellikle hazır yiyecekleri tercih eder ve göbeğinde küçük bir piercing vardır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Gökçe on beş yaşındadır.
2. Gökçe elli yaşındadır.

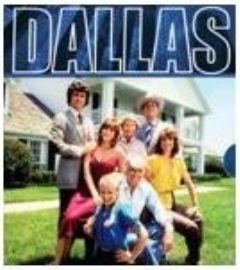
Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i mühendis iken and 995'i avukat idi. Jack is a randomly chosen participant of this study. Berk bu çalışmadan rastgele seçilen bir katılımcıdır.

Berk 36 yaşındadır. O, bekardır ve bir miktar içine kapanıktır. Boş zamanlarını bilim-kurgu okuyarak ve bilgisayar programları yazarak geçirmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Berk mühendistir.
2. Berk avukattır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ünün favori televizyon dizisinin Star Trek, 996'sının ise Dallas olduğu ortaya çıktı. Özgür bu çalışmadan rastgele seçilen bir katılımcıdır.



Özgür 26 yaşında, fizik alanında lisansüstü eğitimine devam eden bir adamdır. O, vaktinin çoğunu evde oturup video oyunları oynayarak geçirmeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Özgür'ün favori televizyon dizisi Star Trek'tir.
2. Özgür'ün favori televizyon dizisi Dallas'tır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si hemşire iken 3'ü doktordu. Yağız bu çalışmadan rastgele seçilen bir katılımcıdır.

Yağız 34 yaşındadır. Lüks bir banliyöde güzel bir evde yaşamaktadır. O, hoşsohbettir ve politikayla yakından ilgilidir. Vaktinin çoğunu kariyerine yatırım yaparak geçirir.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Yağız hemşiredir.

2. Yağız doktordur.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i on altı yaşında iken 995'i elli yaşındaydı. Elif bu çalışmadan rastgele seçilen bir katılımcıdır.

Elif, hip hop ve rap tarzlarında müzik dinlemeyi sever. Genellikle kısa etekler giyer ve arkadaşlarıyla sabaha kadar dans edebileceği partilere katılır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Elif 16 yaşındadır.

2. Elif 50 yaşındadır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü saksafon, 996'sı ise davul çalmaktadır. Toprak bu çalışmadan rastgele seçilen bir katılımcıdır.

Toprak 19 yaşındadır. İstanbul'da okumaktadır ve kız arkadaşı yoktur. Biriktirdiği parayla eski, ikinci el bir araba almıştır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Toprak saksafon çalmaktadır.

2. Toprak davul çalmaktadır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 5'i İstanbul Teknik Üniversitesi mezunu iken, 995'i Boğaziçi Üniversitesi mezunu idi. Tolga bu çalışmadan rastgele seçilen bir katılımcıdır.

Tolga 1.73 boyunda, esmer, iki küçük kız çocuğuna sahip bir babadır. O, üzeri tamamen posterlerle kaplı sarı bir karavan kullanmaktadır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Tolga İstanbul Teknik Üniversitesi mezunudur.
2. Tolga Boğaziçi Üniversitesi mezunudur.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si biyoloji okurken, 3 kişi de kimya bölümünde okumaktadır. Yiğit bu çalışmadan rastgele seçilen bir katılımcıdır.

Yiğit 20 yaşındadır. İstanbul'da okumaktadır ve kalıcı bir kız arkadaşı yoktur. Biriktirdiği parayla henüz kendine ikinci el bir gitar alabilmiştir.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Yiğit biyoloji bölümünde okumaktadır.
2. Yiğit kimya bölümünde okumaktadır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 997'si bilardo oyuncusu iken 3 kişi basketbol oyuncusuydu. Can bu çalışmadan rastgele seçilen bir katılımcıdır.

Can 29 yaşındadır ve hayatı boyunca New York'ta yaşamıştır. O, siyah saçlara ve yeşil renkte gözlere sahiptir. Açık gri renkte bir araba kullanır.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Can bilardo oyuncusudur.
2. Can basketbol oyuncusudur.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 4'ü İzmir'de, 996'sı ise Ankara'da yaşamaktadır. Ayhan bu çalışmadan rastgele seçilen bir katılımcıdır.

Ayhan 28 yaşındadır. O, bir arkadaşıyla apartman dairesini paylaşmaktadır ve bir kız arkadaşı vardır. Basketbol izlemeyi sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. Ayhan İzmir'de yaşamaktadır.
2. Ayhan Ankara'da yaşamaktadır.

Bir çalışmada 1000 kişiye test uygulandı. Katılımcılardan 3'ü yüksek lisansına Yale Üniversitesi'nde, 997'si ise Princeton'da devam ediyordu. John bu çalışmadan rastgele seçilen bir katılımcıdır.

John 22 yaşındadır. Anne ve babası küçük kız kardeşi ile birlikte Toronto'da yaşamaktadır. John zamanının çoğunu kütüphanede klasik eserleri okuyarak geçirir. Bisiklet sürmeyi çok sever.

Hangi seçeneğin olma olasılığı daha fazladır?

1. John yüksek lisansına Yale Üniversitesi'nde devam etmektedir.
2. John yüksek lisansına Princeton'da devam etmektedir.

References for the Images in the Pilot Study

Bogazici University's Logo [Online image]. Retrieved April 28, 2013, from http://tr.wikipedia.org/wiki/Dosya:Bo%C4%9Fazi%C3%A7i_%C3%9Cniversitesi_Logosu.jpg.

Images of a Fifteen Year Old and a Fifty Year Old Woman [Online image]. Retrieved April 28, 2013, from http://www.physiopilatesonline.co.uk/blog/ipilates-to-launch-a-new-way-to-exercise-at-home/468/.

Images of a Nurse, a Doctor, an Engineer and a Lawyer [Online image]. Retrieved April 28, 2013, from http://www.123rf.com/clipart-vector/occupation.html.

Image of a Person Studying Biology [Online image]. Retrieved April 28, 2013, from http://biology.phillipmartin.info/.

Image of a Person Studying Chemistry [Online image]. Retrieved April 28, 2013, from http://www.chemistry.phillipmartin.info/.

Image of a Tattoed Arm [Online image]. Retrieved April 28, 2013, from http://www.masterofink.com/tribal-arm-tattoos/.

Image of a Male and a Female [Online image]. Retrieved April 28, 2013, from http://www.canstockphoto.com/male-female-8694538.html.

Image of a BMW [Online image]. Retrieved April 28, 2013, from http://antalyabmwservisi.com/.

Image of a Renault [Online image]. Retrieved April 28, 2013, from http://www.otobankasi.com/otomobil/renault/renault-renault-12.

Image of Ankara [Online image]. Retrieved April 28, 2013, from http://www.idpieltsturkey.com/tr/ielts-sinavlarimiz-ankarada-sunulmaktadir/.

Image of Izmir [Online image]. Retrieved April 28, 2013, from http://haber.sol.org.tr/kultur-sanat.

Image of Sweden's Flag [Online image]. Retrieved April 28, 2013, from http://tr.wikipedia.org/wiki/%C4%B0sve%C3%A7.

Image of Italy's Flag [Online image]. Retrieved April 28, 2013, from http://tr.wikipedia.org/wiki/%C4%B0talya.

Image of Billiards [Online image]. Retrieved April 28, 2013, from http://www.resimler.co/bilardo/.

Image of Basketball [Online image]. Retrieved April 28, 2013, from http://www.turkiyegazetesi.com.tr/spor/109122.aspx.

Image of Star Trek [Online image]. Retrieved April 28, 2013, from http://www.glogster.com/hannhmuller/izmir/g-6n3qaukc0i8sgisv4o01ba0.

Image of Dallas [Online image]. Retrieved April 28, 2013, from http://tr.wikipedia.org/wiki/Dallas_(dizi,_1978).

Image of a Saxophone [Online image]. Retrieved April 28, 2013, from http://www.saksafon.org/pearl-river-mk007-alto-saksafon/.

Image of a Drum [Online image]. Retrieved April 28, 2013, from http://www.majormusiccenter.com/Vurmalilar/yamaha/Products.aspx.htm.

Image of a Local Market [Online image]. Retrieved April 28, 2013, from http://shintarodan.blogspot.com.tr/2010/10/shopping-at-wang-lung-local-street.html.

Image of a High Quality Store [Online image]. Retrieved April 28, 2013, from http://ariztech.com/digital-promotion-plan-for-fashion-store/.

Istanbul Technical University's Logo [Online image]. Retrieved April 28, 2013, from http://www.bb.itu.edu.tr/tr/genel-bilgiler/kurumsal-ogeler/itu-logo-turkce/image_view_fullscreen.

Kanal D's Logo [Online image]. Retrieved April 28, 2013, from http://tr.wikipedia.org/wiki/Dosya:Kanal_D_logo.png.

National Geographic Channel's Logo [Online image]. Retrieved April 28, 2013, from http://logos.wikia.com/wiki/National_Geographic_Channel_(International).

APPENDIX C: TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü	
Sosyal Bilimler Enstitüsü	
Uygulamalı Matematik Enstitüsü	
Enformatik Enstitüsü	X
Deniz Bilimleri Enstitüsü	

YAZARIN

Soyadı: VERİM
Adı: BURCU
Bölümü: BİLİŞSEL BİLİMLER

TEZİN ADI (İngilizce) : THE IMPACT OF MODALITY AND FEEDBACK ON REASONING ABOUT BASE RATE NEGLECT PROBLEMS IN BEHAVIORAL AND EYE TRACKING STUDIES: A COGNITIVE SCIENCE PERSPECTIVE

TEZİN TÜRÜ: Yüksek Lisans [X] Doktora [ ]

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. [X]

2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) [ ]

3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) [ ]

Yazarın imzası : Tarih : 03.09.2014