THE EFFECT OF CRITERION SHIFTS ON FAMILIARITY AND RECOLLECTION IN RECOGNITION MEMORY

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

BY

ZEYNEP ERBAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF PSYCHOLOGY

FEBRUARY 2022

Approval of the thesis:

THE EFFECT OF CRITERION SHIFTS ON FAMILIARITY AND RECOLLECTION IN RECOGNITION MEMORY

submitted by **ZEYNEP ERBAŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Psychology, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Yaşar KONDAKÇI
Dean
Graduate School of Social Sciences _____

Prof. Dr. Sibel KAZAK BERUMENT
Head of Department
Department of Psychology _____

Assoc. Prof. Dr. Aslı KILIÇ ÖZHAN
Supervisor
Department of Psychology _____

Examining Committee Members:

Dr. Dicle DÖVENCİOĞLU (Head of the Examining Committee)
Middle East Technical University
Department of Psychology _____

Assoc. Prof. Dr. Aslı KILIÇ ÖZHAN (Supervisor)
Middle East Technical University
Department of Psychology _____

Assist. Prof. Dr. Miri BESKEN
İhsan Doğramacı Bilkent University
Department of Psychology _____

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

Name, Last Name: Zeynep Erbaş

Signature:

ABSTRACT

THE EFFECT OF CRITERION SHIFTS ON FAMILIARITY AND RECOLLECTION IN RECOGNITION MEMORY

ERBAŞ, Zeynep
M.S., The Department of Psychology
Supervisor: Assoc. Prof. Dr. Aslı KILIÇ ÖZHAN

February 2022, 102 pages

Familiarity and recollection refer to memory processes that are engaged during recognition judgments. A task that is widely used to operationalize these processes is the remember-know task. In this thesis, it was aimed to examine the effect of different decision criteria on the corresponding "remember" and "know" responses with respect to the signal detection model of recognition memory. In line with this purpose, a standard recognition memory task was applied with a between-subjects criterion manipulation. A conservative response bias was induced by presenting participants a test list, in which only 30% of the items were previously studied. Conversely, a liberal response bias was induced by presenting participants a test, in which only 70% of the items were previously studied. The results from two experiments showed that the base rate manipulation led to a shift in response criterion, which was calculated based on the assumptions of the signal detection theory. Moreover, "know" and "remember" responses were associated with responses made with lower and higher confidence, as well as responses that were associated with lower and higher response thresholds, respectively. Hence, the obtained results are in line with the assumptions of the

single-process models of recognition memory based on the finding that the two processes can be represented on a mutual medium (i.e. memory strength). However, the lack of an effect in terms of the corresponding hit and false alarm rates for “remember” and “know” responses require further research in order to distinguish or associate the two processes.

Keywords: Recognition Memory, Response Bias, Criterion, Signal Detection Theory

ÖZ

TANIMA BELLEĞİNDE KRİTER DEĞİŞİMİNİN AŞİNALIK VE HATIRLAMA ÜZERİNDEKİ ETKİSİ

ERBAŞ, Zeynep
Yüksek Lisans, Psikoloji Bölümü
Tez Yöneticisi: Doç. Dr. Aslı KILIÇ ÖZHAN

Şubat 2022, 102 sayfa

Aşinalık ve hatırlama, tanıma kararları sırasında devreye giren hafıza süreçleridir. Bu süreçleri ölçülebilir hale getirmek için yaygın olarak kullanılan bir görev, hatırlama-bilme görevidir. Bu tezde, farklı karar kriterlerinin "hatırlama" ve "bilme" yanıtları üzerindeki etkisinin tanıma belleğinin sinyal tespit modeline göre incelenmesi amaçlanmıştır. Bu amaç doğrultusunda standart bir tanıma belleği görevi içerisinde kişiler arası bir kriter manipülasyonu uygulanmıştır. Bu doğrultuda katılımcılara test maddelerinin sadece %30'unun daha önce çalışıldığı bir test listesi sunularak muhafazakar bir yanıt yanlılığı oluşturulmuştur. Bunun aksine katılımcılara maddelerin sadece %70'inin daha önce çalışıldığı bir test listesi sunularak ise liberal bir yanıt yanlılığı oluşturulmuştur. İki deneyden elde edilen sonuçlar, temel oran manipülasyonunun, sinyal tespit teorisinin varsayımlarına dayalı olarak hesaplanan yanıt kriterinde bir kaymaya yol açtığını göstermiştir. Buna ek olarak "hatırlama" ve "bilme" yanıtları, sırasıyla daha düşük ve daha yüksek güvenle verilen yanıtlarla ve ayrıca sırasıyla daha düşük ve daha yüksek yanıt eşikleri ile ilişkilendirilmiştir. Bu nedenle, elde edilen sonuçlar iki sürecin ortak bir payda ile (yani hafıza gücü) temsil

edilebileceği bulgusuna dayanan tek süreç tanıma belleği modellerinin varsayımları ile uyumludur. Bununla birlikte, “hatırlama” ve “bilme” yanıtlarına karşılık gelen isabet ve yanlış alarm oranları açısından bir etkinin bulunmaması, iki süreci ayırt etmek veya ilişkilendirmek için daha fazla araştırmanın gerekliliğini vurgulamaktadır.

Anahtar Kelimeler: Tanıma Belleği, Cevap Yanlılığı, Kriter, Sinyal Tespit Teorisi

To my grandfather Yaşar Denli, with whom I shared my first sparks of excitement for science

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor Assoc. Prof. Dr. Aslı Kılıç Özhan for her guidance and support throughout my graduate studies. She always approached me with the utmost understanding and patience during a time which coincided with the pandemic period, and that, I will remember even years from now. I have matured as a person and as a scientist under her supervision.

I would like to thank Dr. Dicle Dövencioğlu and Assist. Prof. Dr. Miri Besken for accepting to be a member of the examining committee of my thesis. Their time and suggestions were invaluable in helping this thesis take its final form.

I owe many thanks to the members of the METU Memory Modeling Lab for all the discussions we had together, and the friendships they gave me. We have all gone through an unknown and stressful time during this pandemic period in our personal and academic lives, and knowing that we were sharing similar struggles has made me feel less alone.

I am thankful to my family and friends for being in my life. I am grateful to have a mother who is my biggest cheerleader, and a brother to whom I can always count on to give me the harshest criticisms! They have a way of lifting me up and humbling me all at once. I would also like to thank my friends for being with me on this journey and standing by me through all the ups and downs.

Finally, I would like to thank my fiancé Kemal Yardımcı for his never ending love and support. I am grateful every day for having met you before all hell broke loose. You were, and still are, always there to help me get back on my feet. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Recognition Memory

The idea that human memory is composed of multiple systems has been proposed by many philosophers and psychologists in the past (James, 1890; Ryle, 1949; Herrmann, Douglas, & Chaffin, 1988). Despite the valued contributions of philosophical discourse on the various taxonomies of memory, a more objective and parsimonious classification was proved to be necessary due to the lack of a unified view on the matter. The acceleration of experimental inquiries regarding the classification of memory structures has brought the field closer to presenting a single view on how the information is stored in the brain. In these inquiries, most of the arguments proposing distinct memory systems were based on evidence from healthy participants (Tulving, Schacter, & Stark, 1982; Tulving & Schacter, 1990), amnesic patients (Milner, 1962; Milner, Corkin, & Teuber, 1968; Warrington & Weiskrantz, 1970; Shimamura & Squire, 1987), and experimental lesion studies with animals (Schmaltz & Theios, 1972; Packard, Hirsh, & White, 1989). Accumulating evidence from these studies suggested the presence of two distinct memory systems: one that involves active recollection and one that involves memory for performance.

The taxonomy developed by Squire (1987, 2004) divides memory into declarative and nondeclarative memory; the former refers to representations of events that can be consciously retrieved via recollection, and the latter refers to performance related memories that are retrieved via the activation of previously learned memory traces.

Declarative memory is further divided into episodic and semantic memory (Tulving, 1984, 1993). The episodic memory constitutes memory for events, and the semantic memory constitutes memory for facts, without the contextual details. Episodic and semantic memory can be examined with particular memory tasks that are assumed to tap into certain memory systems. In particular, episodic memory is widely investigated in laboratory environments with recognition memory tasks.

In attempts to examine which memory system taps into particular memory tasks, recognition memory paradigms are implemented in order to differentiate the ways in which people engage in retrieval processes. In a standard recognition memory task, item recognition is tested with participants studying a list of items (words, pictures, etc.), and later being tested on them. The study list consists of a number of items that the participants are asked to study. There is usually a brief distractor between the study and test phases in order to prevent any instances of rehearsal, as well as any recency effect (i.e. a serial position effect in which the most recent items have a greater chance of later being retrieved; Ebbinghaus, 1913/1885; Murre & Dros, 2015). The test list consists of targets (previously studied items) and foils (new items). The participants' task at test is to discriminate the targets from the foils. The memory performance is measured by the proportion of old or yes responses to target items (hit rate) and the proportion of old or yes responses to foils (false alarm rate).

1.2. Signal Detection Theory

The standard model of Signal Detection Theory (SDT; Green & Swets, 1966; Banks, 1970) is a mathematical model for discriminating between two sources of information (See Figure 1.1). These two sources are signal and noise. In the human memory application of the SDT model, the memory traces are defined as the signal, which is often referred to as the familiarity value. Accordingly, the target items constitute the signal component of the model whereas the foils constitute the noise component. In this model, the signal and noise distributions are defined as

two normally distributed graphs. The distance between the signal and noise distributions determines the discriminability of signal and noise. The closer the graphs, the more the signal and noise information are intermingled, and the more difficult it is to discriminate them. The further away the graphs from one another, the more easily signal and noise information are separated, and it is easier to discriminate them. For instance, an easier task (e.g. a task where items are studied three times and are easily retrieved from memory) may be illustrated by distributions that are distant whereas a more difficult task (e.g. a task where study items are only flashed briefly) may be illustrated by distributions that are close to one another.

According to SDT models, old/new decisions are based on a recognition confidence continuum that can be mapped onto the memory strength distributions for old and new items. These two distributions are assumed to be normally distributed, and the mean memory strength for the old item distribution is greater than the mean memory strength for the new item distribution. During decision making, participants set a response criterion somewhere on these distributions, and a recognition decision is made based on whether the strength of a memory item exceeds the response criterion or not. If the memory strength of an item exceeds the response criterion, an "old" response is given; if the memory strength of the item is below the response criterion, a "new" response is given.

In addition to the distribution characteristics, the memory judgment is ultimately determined by the criterion. The criterion is a value set by participants at which the items whose familiarity exceeds the criterion are identified as old, and the items whose familiarity is below the criterion are defined as new. Hence, the placement of the criterion value is one of the factors that affect memory judgments. A more liberal criterion, in which the criterion line is drawn to the left of the neutral position, results in more HR and more FAR. In contrast, a more conservative

criterion, in which the criterion line is drawn to the right of the neutral position, results in less HR and less FAR (See Figure 1.1).

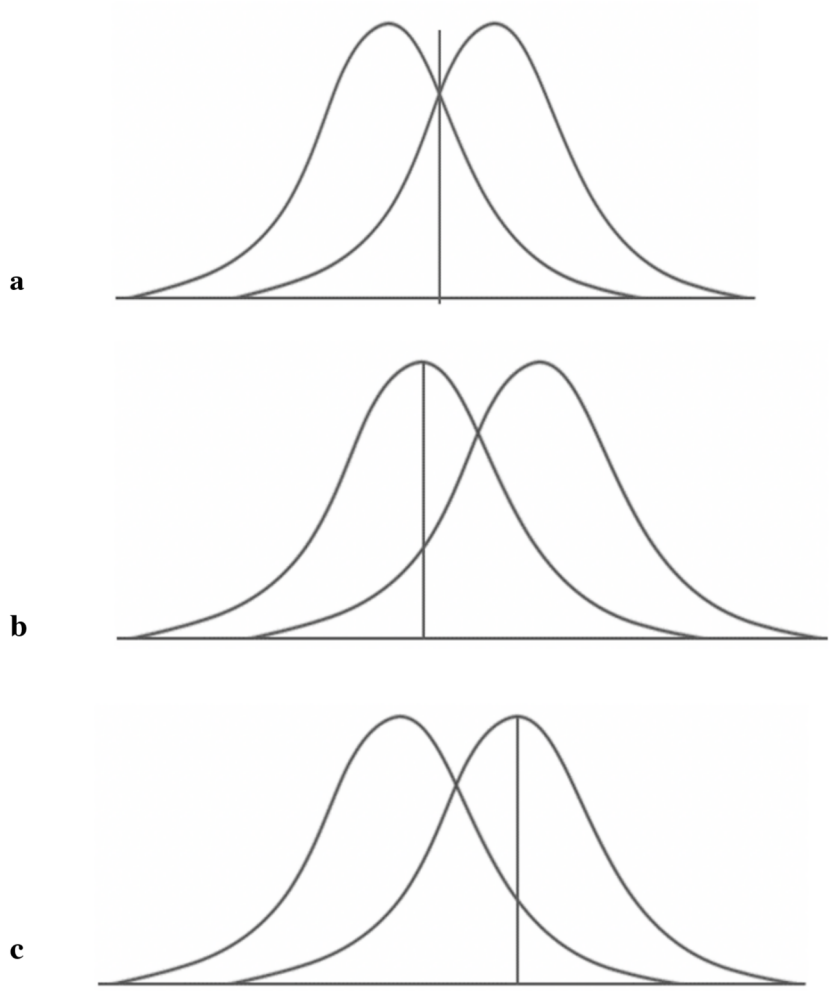

Figure 1.1 An illustration of criterion shifts on the SDT model. The neutral (a), liberal (b) and the conservative (c) criteria are illustrated with vertical lines.

In this context, if a target item is correctly identified as old, this is defined as a hit (H). If the target item is falsely identified as new, then this is defined as a miss (M). Alternatively, identifying a foil as old is defined as a false alarm (FA), and

identifying a foil as new is defined as a correct rejection (CR; see Table 1.1 and Figure 1.2). The memory performance is defined in terms of the discrimination ability between signal and noise information, that is, the proportion of hit rate (HR) and false alarm rate (FAR).

Table 1.1
Four Outcomes of Signal Detection Theory

Item / Response	Yes	No
Target	H	M
Foil	FA	CR

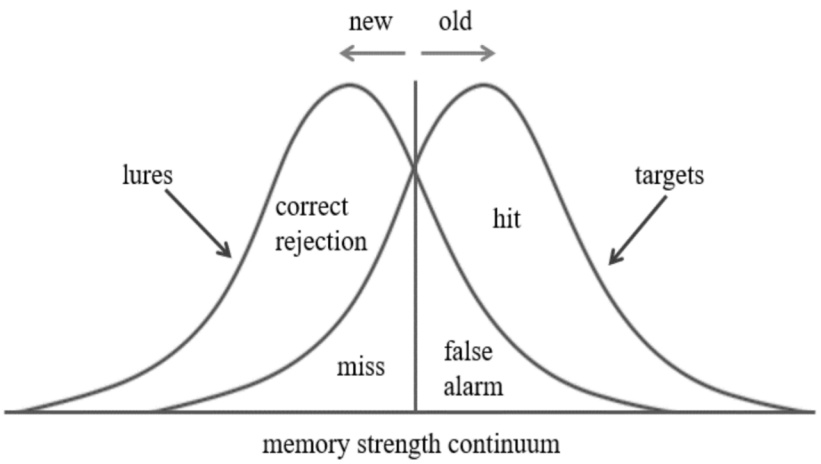

Figure 1.2 A distribution depicting the basic principles of the Signal Detection Theory.

In addition to measuring memory performance, test manipulations affecting memory strength and decisional factors have been applied on recognition memory tasks in order to understand the underlying dynamics of memory (Hirshman, 1995;

Murdock & Dufty, 1972; Stretch & Wixted, 1998). In particular, the application of the SDT model to standard recognition memory tasks is a widely used paradigm to examine strength and criterion related components in recognition memory judgments (Donaldson, 1996a; Donaldson, 1996b; Macmillan & Creelman, 2005; Wixted, 2007; Wixted & Stretch, 2004). In such applications, model parameters are calculated and/or adjusted in order to account for the effects of interest. In this regard, the sensitivity parameter, represented by *d'* (dee-prime; Macmillan & Creelman, 2005), indicates how well items can be discriminated. It also indicates the distance between the signal and noise distributions. The *d'* value is calculated in terms of the z scores of HR and FAR as follows (Macmillan & Creelman, 2005, p. 8):

(1) $$d' = z(HR) - z(FAR)$$

The z calculation converts HR and FAR values into standard deviation units so that a proportion of 0.5 (i.e. when the proportion of H and FAs are equal) is converted into a z score of 0. Smaller proportions are converted into negative z scores, and larger proportions are converted into positive z scores. Hence, sensitivity increases as *d'* increases. An important note referring to *d'* calculations is that there are normality and equal variance assumptions for the signal and noise distributions. In case that these assumptions are not met, non-parametric calculations are applied (Macmillan & Creelman, 1996).

The bias parameter, represented by C, indicates the placement of the subjective criterion, and the distance from the point where the signal and noise distributions intersect (Macmillan & Creelman, 2005). The C value is calculated as below:

(2) $$C = -0.5[z(HR) - z(FAR)]$$

Accordingly, a *C* value of 0 indicates a neutral criterion, and a criterion line that cuts the intersection point between the signal and noise distributions. A negative *C* score indicates a lenient or liberal criterion whereas a positive *C* score indicates a strict or conservative criterion.

1.3. Single- and Dual-process Models of Recognition Memory

1.3.1. Single-process Models

In explaining the findings from recognition memory experiments, two model accounts have been proposed predominantly: Single-process models and dual-process models. Among these models, single-process models suggest that item recognition decisions are based on a single process that is affected by memory strength and decisional factors (Donaldson, 1996; Wixted & Stretch, 2004; Wixted, 2007). In that sense, the single process models are based on the SDT interpretation of recognition memory such that any dual, or multiple, memory judgments that are beyond the old/new distinction can be explained on the basis of different thresholds on the memory strength continuum. Moreover, according to the single process models of recognition memory, the target distribution is wider than the lure distribution, resulting in the unequal variance signal detection model. For example, in a study examining the SDT model fits of data collected for a developmental visual recognition memory experiment, the distribution characteristics of single and dual process models were examined (Hayes, Dunn, Joubert & Taylor, 2016). When the dual-process high threshold signal detection model, and several single-process models (equal variance signal detection, unequal variance signal detection, mixture signal detection) were fit to obtained data, the mixture and unequal variance signal detection models resulted in better fits than the equal variance model. In another study, it was found that the standard deviation of confidence ratings made to target items was larger than the standard deviation of confidence ratings made to lures in a recognition memory experiment

(Mickes, Wixted & Wais, 20017). The results further suggested a wider target distribution as compared to the lure distribution. The unequal variance signal detection model has been further shown to provide a better fit of the old/new recognition memory data in several other studies (Slotnick, Klein, Dodson & Shimamura, 2000; Spanton &. Berry, 2020; Yonelinas, 1999).

1.3.2. Dual-process Models

An alternative to the single-process models, dual-process models suggest that item recognition decisions are based on two qualitatively distinct memory processes (Mandler, 1980). According to these models, familiarity and recollection processes both contribute to the decision as to whether an item is recognized or not. The best example for these two processes comes from Mandler's (1980) "butcher-on-the-bus" anecdote. In this thought experiment, one sees a familiar face on the bus and tries to search his/her memory to find the source of this feeling. The experience is so strong that the person is confident that this face is of someone he/she encountered before; however, he/she can't remember who this person is or where he/she saw that face before. Eventually the person realizes that the face belongs to the butcher from the supermarket. The feeling of knowing up until the moment the person remembers where he/she knows this person from represents familiarity; the later realized retrieval of the information along with the contextual details represents recollection. This example dissociates the two processes proposed to underlie the dual processes of recognition memory. Accordingly, familiarity refers to the automatic, subjective feeling of knowing, and recollection refers to the controlled retrieval of both the item information and the context attached to the information. Proponents of the dual-process model argue that the two components of recognition memory, familiarity and recollection, represent two qualitatively different processes. In particular, it is argued that familiarity is a continuous process that corresponds to the memory strength of an item whereas recollection is a threshold process that occurs in an all-or-none manner (Onyper, Zhang & Howard, 2010; Yonelinas, 1994). This is in contrast to the

interpretation of the familiarity and recollection judgments according to the single process model assumption, which argues that both familiarity and recollection judgments are continuous processes.

1.4. Remember-Know Paradigm

There is an ongoing debate as to which of the models best explains the recognition memory process. Several behavioral studies have been conducted so far, in attempts to favor one model over the other. In favor of the dual-process models, a procedure known as the remember/know (or RK) paradigm has been applied extensively in the literature. The procedure was originally proposed by Tulving (1985) to measure different states of awareness that underlie memory retrieval. In attempts to dissociate which memory system is engaged in a particular memory task, Tulving, Gardiner and like-minded researchers (See Gardiner & Java, 1990, 1993; Tulving, 1985) applied the RK paradigm to examine participants' self-reports on how they decide whether an item has been studied. In order to guide participants' responses in accordance with the presumed framework, the procedure required participants to make a *remember* judgment if the test item is retrieved from episodic memory, and a *know* judgment if the test item is retrieved from semantic memory. In particular, participants were instructed to respond "remember" (or R) if they remembered the test item and the contextual details, and respond "know" (or K) if they know that the test item was presented before but they fail to remember any contextual details along with it. More recently, the method has been used as a measure of familiarity and recollection in episodic memory.

The R and K judgments were interpreted to denote a state of autonoetic awareness and noetic awareness, respectively. In this context, autonoetic awareness was suggested to refer to retrieval from episodic memory whereas noetic awareness

was suggested to refer to retrieval from semantic memory. The RK paradigm was operationalized in terms of more functional terms by Gardiner (1988) who defined *remembering* as a conscious recollection of an item, and *knowing* as "recognition memory in the absence of conscious recollection" (p. 310). Gardiner (1988) further examined the relation between R and K judgments, and whether the two responses reflect functionally separate memory systems. He found that the proportion of R responses was affected by levels of processing manipulations (Craik & Lockhart, 1972; Craik & Tuving, 1975) whereas K responses were unaffected by the same manipulation. That is to say that among R responses, items that were encoded with a semantic orienting task (i.e. a task where participants are asked to process the word item and come up with semantically associated items, which strengthens the encoding of the item and increases chances of later retrieval) were recognized more successfully as compared to words, of which only a rhyming task was given. However, among K responses, there was no difference in the recognition performance between words that were studied deeply and words that were studied shallowly. The results were suggested to be consistent with the assumptions of dual-process models of recognition as the R and K responses were functionally dissociated under the levels of processing manipulation.

Many findings obtained with the RK paradigm suggest the existence of a dissociation between the two distinct processes (Gardiner, Gregg, & Karayianni, 2006; Gardiner, Ramponi, & Richardson-Klavehn, 1998; Gardiner & Richardson-Klavehn, 2000; Knott & Dewhurst, 2007; Yonelinas et al., 1998). In this context, a widely recognized model of the RK paradigm is by Jacoby, Yonelinas, and Jennings (1997) who suggest that R and K judgments reflect recollection and familiarity processes in recognition memory, respectively (for a review, see; Yonelinas, 2002). Therefore, in studies using the RK procedure, R responses indicate recollection, and K responses indicate familiarity. In support of the dual-process theory, there is extensive behavioral evidence from double dissociations between R and K responses on the basis of recognition tasks (Schacter & Tulving,

1994; Yonelinas et al., 1998), the features of the receiver operating characteristic (ROC) curve (Yonelinas, 1999), and processing speed differences (McElree, Dolan, & Jacoby, 1999) as well as neurological evidence from lesion studies (Bastin et al., 2004; Bowles et al., 2007) and electrophysiological recordings (Düzel et al., 1997; Rugg et al.,1998). However, these findings have been challenged by single-process models, which suggest that familiarity and recollection judgments reflect gradual levels of confidence on a memory strength continuum (Wixted, 2007). Hence, it is argued that any line of evidence pointing to a functional dissociation may be explained with the SDT model by changing the respective parameters of the model in line with the proposed manipulations. According to the SDT interpretation of the RK paradigm, familiarity judgments are based on a lower threshold than recollection judgments (Donaldson, 1996). In line with this view, it was found that recollection judgments have higher confidence levels than familiarity judgments (Stretch & Wixted, 1998; Wixted & Stretch, 2004). It was further shown that even the false alarms for R responses were given with higher confidence than the confidence ratings of hits for K responses. This suggests that R responses are given with higher confidence as compared to K responses regardless of response accuracy, and that confidence levels modulate the transition between the two processes.

1.4.1. Donaldson's Model

The model proposed by Donaldson (1996) argues for a two-threshold model of signal detection, in which the first threshold determines whether an item is judged as old or new, and the second threshold determines whether an item is known or remembered. Hence, there are three possible responses according to this model. First, if an item falls below the first threshold, it is identified as new. Second, if it falls between the two thresholds, it is identified as old and is given a K judgment. And finally, if it falls above the second criterion, it is identified as old and is given

a R judgment. There are situations in which the placement of these criterions may change with respect to decision biases. For instance, a liberal criterion may be adopted, as compared to a neutral criterion, which implies more likelihood of giving an *old* response. Alternatively, a conservative criterion may be adopted, which implies more likelihood of giving a *new* response (See Figure 1.3).

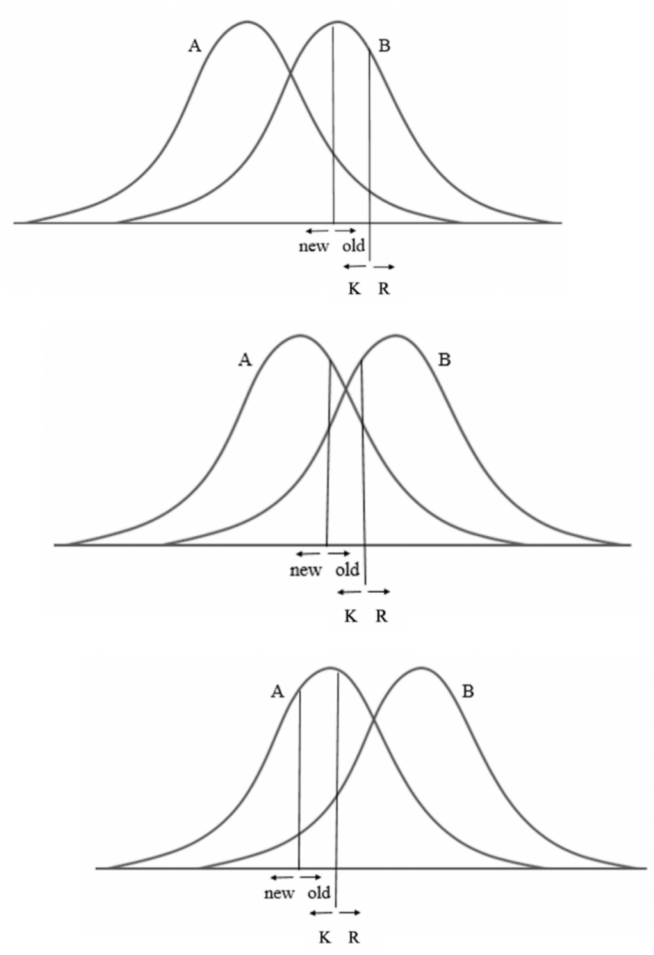

Figure 1.3 An illustration of Donaldson's (1996) conceptualization of RK judgments within the signal detection framework. The upper panel represents the conservative condition, the middle panel represents the neutral condition, and the bottom panel is the liberal condition.

A consideration of the three possible placements of criteria reveals predictions about the proportion of old/new and R-K responses, and performance estimates as well as respective HRs (hit rates) and FAs (false alarm rates). When the distributions in Figure 1.3 are examined, the area under Distribution A between two criteria represents FAs, and the area under Distribution B between two criteria represents Hs for K responses. Similarly, as for the items to the right of the second criterion, the area under Distribution A represents FAs, and the area under Distribution B represents Hs for R responses. As a result, the criterion on the right represents the subdivision of the old responses into R and K responses. In line with these descriptions, the model puts forth some predictions about the RK paradigm.

First, (1) the sensitivity parameter (i.e. *d'*) provides a criterion-free estimate of memory performance; therefore, *d'* for R responses should be equal to *d'* for overall recognition performance. Second, (2) considering the three possible criterion conditions (i.e. liberal, neutral and conservative criteria), K responses are not independent from the old/new response threshold under differing criteria. Accordingly, the HR should be higher than the FAR among K responses under conservative criteria, equal under neutral criteria, and lower under liberal criteria. In addition, *d'* takes its largest value under conservative criteria, equals to 0 under neutral criteria, and is less than 0 under liberal criteria. The rationale behind these predictions can be observed from the areas under distributions A and B for the three respective conditions. When responding is conservative, the area under Distribution B between the two thresholds is larger than the area under Distribution A between the two thresholds, which corresponds to a higher HR than FAR. Consequently, the calculation of *d'* will be larger when HR is higher than FAR. As responding becomes more neutral, the HR and FAR will be about equal to one another, that is, the areas under distributions A and B between the two thresholds will be equal, and performance will be at chance level. Finally, as responding gets more liberal, the area under Distribution B is smaller than the area under Distribution A between the two thresholds. Thus, the FAR is higher than HR, *d'* is negative, and the performance is below chance levels. This suggests that there is a

positive correlation between the placement of the old/new criterion and memory performance for K responses: K response performance gets better as the criterion becomes more conservative. Third, (3) any effect of a manipulation on the overall memory performance will be parallel to its effect on R response performance.

Here is how Donaldson (1996) explains this: If we were to imagine a third distribution to the left of Distribution A, and called it the new item distribution, distributions A and B would be old item distributions. Then, due to the larger overlap between the new distribution and Distribution A, Distribution A would represent a more difficult level of a manipulation while due to the smaller overlap between the new distribution and Distribution B, Distribution B would represent an easier level of a manipulation (See Figure 1.5). In this scenario, the HR will always be higher for the easy task than it is for the more difficult task. Regardless of where the criterion moves, above the second threshold, the area under Distribution B (i.e. HR for R responses) will always be larger than the area under Distribution A (i.e. FAR for R responses). Hence, a variable that affects overall memory performance will affect R responses in the same direction. Finally, as for the effect of a variable on K responses, the following pattern is predicted: When responding is conservative, the R and K responses will show the same effect. When responding is neutral, although the variable has an effect on R responses, it will have no effect on K responses. When responding is liberal, the effect of the variable on K responses will be in the opposite direction to its effect on R responses (See Table 1.2). These predictions are obtained from the relative areas under distributions A and B between the two thresholds, that is, the relative values of HR and FAR for K responses.

Table 1.2

Predictions of Donaldson's (1996) Model for Possible Outcomes of R/K Responses Under Different Criteria

Criterion / Response	Know	Remember	Overall
Conservative	decrease / increase	decrease / increase	decrease/increase
Neutral	no effect	decrease / increase	decrease/increase
Liberal	decrease / increase	increase / decrease	increase/decrease

1.5. Criterion Shift

The criterion shifts in recognition memory has been of interest to many researchers (Han & Dobbins, 2008, 2009; Hicks & Starns, 2014; Rotello et al, 2006; Rhodes & Jacoby, 2007). The criterion shifts can be experimentally applied and/or observed through methods such as instructions (Postma, 1999), memory strength manipulations (Hirshman, 1995; Stretch & Wixted, 1998; Verde & Rotello, 2007), base rate manipulations (Cox & Dobbins, 2011; Estes & Maddox, 1995; Healy & Kubovy, 1977, 1978; Rhodes & Jacoby, 2007), payoff manipulations (Curran, Debuse & Leynes, 2007), and probabilistic mnemonic cues (Selmeczy & Dobbins, 2013). In the matter of criterion shifts in recognition memory, there have been arguments put forth by proponents of single-process models as to whether the proposed RK distinction refers to a distinction relating to the decision criterion that is placed on the memory strength continuum (Donaldson, 1996; Hirshman & Master, 1997). In this context, the effect of criterion shifts on the respective RK judgments has been studied extensively (Gardiner, Richardson-Klavehn & Ramponi, 1997; Hirshman & Henzler, 1998; Postma, 1999; Rotello et al., 2006; Strack & Foerster, 1995).

The idea that R and K responses in recognition memory experiments reflect two functionally distinct memory systems was criticized by Strack and Foerster (1995). They argued that participants's responses don't simply reflect the existing memory systems underlying the respective memory tasks but rather, their self-reports are "affected by situational cues that are exogenous to the theories of memory that these reports are meant to test" (Strack & Foerster, 1995, p. 357).. The criticism was mostly based on the study by Gardiner and Java (1990) in which a crossover interaction was found between R and K responses, and old words and nonwords. In their experiment, participants gave more R than K responses for old words, and more K than R responses for old nonwords. In addition, it was found that low-frequency words resulted in more R responses than high-frequency words whereas K responses were unaffected by the word frequency effect. The finding of a crossover interaction and a null effect for K responses strengthened the idea of a functionally dissociable memory system that can be measured by self-reports of R and K judgments. However, these findings were criticized on the basis of the confounding nature of the RK paradigm such that the R and K responses were codependent in the sense that the lack of an R response necessitated a K response, and the very nature of the task was against the independence assumption of the two processes. Moreover, the availability of only two potential responses exhausted any other means by which a recognition memory judgment can be made. From this point of view, Strack and Foerster (1995) argued that the use of judgmental strategies may elicit K responses by default due to the *inferential* experiment designs (See also Rajaram, 1993). They further argued that participants may base their judgments on their prior metacognitive knowledge about whether an item has been studied. Hence, a criterion manipulation has been applied via instructions to induce prior expectations on participants' later response tendencies, and to use a manipulation that is known not to affect memory performance. They found that K responses were more likely to be influenced by response bias manipulations than R responses when participants were told either 50% or 30% of the test items consisted of the studied items. In response to Strack and Foerster (1995), a follow up study by Gardiner, Richardson-Klavehn and Ramponi (1997)

was conducted, in which they instructed participants similarly that either the 30% or the 50% of the test list consisted of studied items. They found that response bias had no effect on either R or K responses when an additional "guessing" option was added to the potential responses, and that only the guessing responses were influenced by the criterion manipulation.

Although the selective effect of criterion manipulations on RK judgments and the crossover interactions supported the existence of a dual-process model, the development of a two-threshold signal detection model (Donaldson, 1996; Hirshman & Master, 1997) has provided an alternative perspective to the RK paradigm. Hence, empirical tests of this paradigm have gained momentum. Arguably the most fundamental prediction of the two-threshold signal detection model is that experimental manipulations that change the placement of the criterion should affect both R and K judgments. The fact that the very nature of the model incorporates factors such as criterion placement and signal strength makes it a great candidate to test the effect of criterion manipulations on the respective R and K judgments. One of the first studies to test this was by Hirshman and Henzler (1998) who argued for a two threshold signal detection model of familiarity and recollection judgments.

1.6. Aims of the Thesis Research

The contrasting lines of evidence in support of both models prevented reaching a conclusion using evidence. All of the studies that are mentioned above defined familiarity as a continuous process. However, there have been debates as to whether recollection is a continuous or a threshold process. The studies that characterized recollection as a threshold process generally assumed that familiarity, as measured by K judgments, reflects a continuous process that can be measured on the basis of memory strength, whereas R judgments are independent of a memory strength continuum, and are made according to whether a certain

threshold has been passed or not (). As a result, K and R judgments are proposed to be dissociable from one another. In contrast, other studies suggested that recollection, too, is a continuous process (). In response to the findings supporting dual-process models, a number of criticisms have been raised by proponents of single-process models. A substantial amount of evidence has been presented which argued that the definition and the of familiarity were confounded by other mediating factors. In this context, memory strength and response criteria have been argued to constitute the common medium on which R and K judgments are laid. These arguments suggest that familiarity and recollection processes can be separated in terms of low and high confidence responses, or responses made on the basis of low and high criteria, respectively.

In the current study, the single-process SDT interpretation of familiarity and recollection is explored in two experiments examining the effects of criterion on the corresponding R and K judgments. In this direction, it is hypothesized that R and K judgments will reflect high and low threshold responses, as well as high and low confidence responses. Moreover, in line with the predictions of Donaldson's model (1996), the following results are hypothesized in terms of hit and false alarm rates, given that the area between the old/new and the remember thresholds represent "know" responses. First, HRs should be higher than the FARs in the conservative criterion, and lower than FARs in the liberal criterion. Second, *d'* should provide a criterion free estimate of memory, meaning that it should be equal under both conditions. However, Donaldson argues that memory performance may be below chance under the liberal criterion as the HR is less than the FAR in the distribution of the liberal condition. Hence, a *d'* value smaller than 0 is still within the expected values according to the model. A pattern of results that are in line with these predictions will support a single-process model interpretation of familiarity and recollection judgments in terms of Donaldson's model. However, note that alternative interpretations of the single-process model remain other than this model considering its strict predictions regarding HRs and FARs in particular.

Conversely, a lack of a relation between criterion or confidence and the corresponding familiarity and recollection judgments will suggest that the two processes are independent of each other, which will support the dual-process account.

CHAPTER 2

EXPERIMENT 1

2.1. Method

The rationale of the study is motivated by the SDT interpretation of recognition memory judgments, which proposes that R-K judgments map onto a memory strength continuum based on their respective levels of confidence (Wixted & Micked), and corresponding hit and false alarm rates depending on the areas under the memory strength distribution given the placement of the criteria for memory judgments (Donaldson, 1996). According to this account, subjective judgments of familiarity and recollection reflect a quantitative distinction rather than a qualitative one. Therefore, studying the old/new judgments of participants in a recognition memory experiment under a criterion manipulation, along with their additional R-K judgments and confidence ratings will provide a data set that is fit to examine the consequences of familiarity and recollection judgments in a single-process model paradigm.

2.1.1. Participants

40 undergraduate students were recruited from Middle East Technical University (METU) to participate in the study. The experiments conducted were approved by the METU Human Subjects Ethics Committee (Appendix A). Students received a partial course credit in return for their participation. All participants provided written consent

prior to the beginning of the experiment (Appendix B). The data from three participants were not complete due to errors related to MATLAB: 1 participant accidentally exited the experiment by pressing the Window key, the number of trials was mistyped for one participant, and there was a MATLAB related error that ended the experiment for one participant. The data from another participant were excluded because the experiment was disrupted by someone entering the lab and talking to the participant during the course of the experiment. Therefore, after removing these four participants, the complete data from 36 participants (M age= 22.4, SD age = 3.75) were used in the analysis. There were 7 males and 29 females, and 3 left-handed and 33 right-handed people. All participants were native Turkish speakers with normal color-vision and normal or corrected-to-normal visual acuity.

2.1.2. Materials

Words were randomly selected from the Turkish Word Norms (Tekcan & Göz, 2005) to be used in the recognition memory experiment. The words used in the experiment ranged from four to nine letter words. A complete list of 1232 words were typed in a text document to be later used in the MATLAB script of the experiment. The experiment was coded in MATLAB using the Psychtoolbox extension.

2.1.3. Design and Procedure

The experiment was a two factor (base rate: fixed or manipulated and criterion: liberal or conservative) between subject design. Participants were randomly assigned to each condition. During the course of the experiment, there was one practice block consisting of 10 study trials and 10 test trials, and four experimental blocks each of which consisted of 80 study trials and 80 test trials. The word list was shuffled before each session, and randomly presented to each participant.

For the base rate manipulation, half of the participants were told that they would be presented with a test list in which 30% of the test items would consist of studied items while the other half was told that they would be presented a test list in which 70% of the test items would consist of studied items, when in fact, a base rate manipulation was made only in half of the participants in each group. The other half was tested on a list that had equal numbers of targets and foils.

The 30% base rate condition constituted the conservative condition, and the 70% base rate condition constituted the liberal condition. Accordingly, participants in the 30% base rate condition were expected to set a stricter criterion and be more likely to respond "new" to the test items. On the other hand, participants in the 70% base rate condition were expected to set a leaner criterion and be more likely to respond "old" to the test items.

Table 2.1
Number of Targets and Foils Presented in Each Condition

Base Rate/Criterion	Conservative	Liberal
Base Rate Manipulation	24 (T) – 56 (L)	56 (T) – 24 (L)
No Base Rate Manipulation	50 (T) – 50 (L)	50 (T) – 50 (L)

Participants provided demographic information in the beginning of the experiment including questions regarding their age, gender and handedness (left or right). They were then asked to read a written instruction, which was repeated once more with the experimenter, and encouraged to ask any further questions before beginning the

experiment. All participants were provided with an example to explain the difference between *remember* and *know* responses to make sure that the distinction was clear. The example provided was based on the original example given by Mandler (1980) but it was changed slightly to be more relatable for students (see Appendix C, for a full version of the instruction text). Participants were not allowed to ask any questions or to interact with the experimenter during the experiment.

In the study phase, participants were instructed to attend and study the words that would appear on the screen one after another. A fixation cross was presented before the presentation of the word stimulus for 400 milliseconds, and each word was presented on the screen for 1 second. The study phase was followed by a distractor task in which participants were asked to make arithmetic calculations for 45 seconds before proceeding to the test phase. The purpose of the distraction phase was to put an interval between the study and test phases, and to prevent participants from rehearsing the studied words. During the distraction phase, participants were instructed to add the number presented on the screen to the current sum until the duration of the distraction task was completed (Figure 2.1).

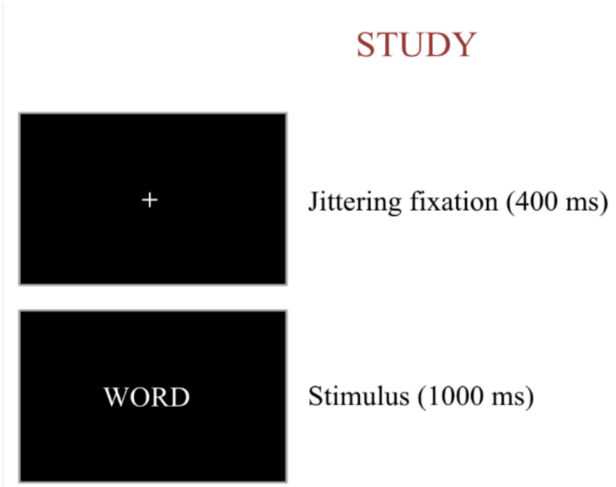

Figure 2.1 An illustration of the study phase and the distractor task.

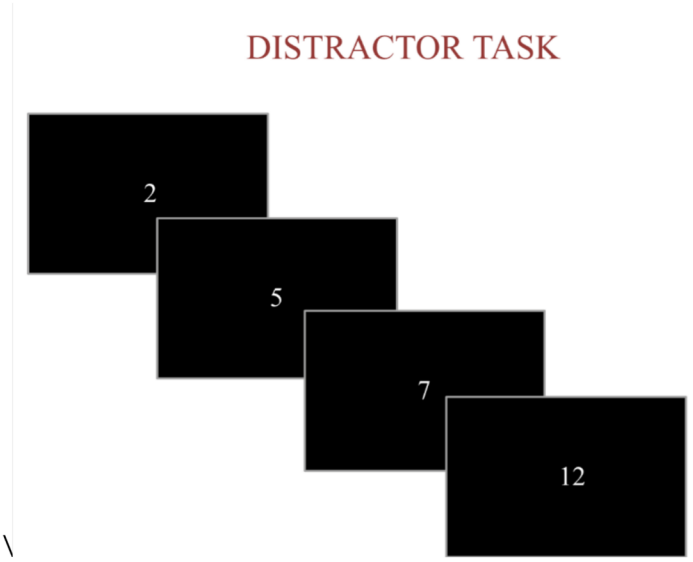

Figure 2.1 (cont'd) An illustration of the study phase and the distractor task.

The test phase began with a jittering fixation, the duration of which ranged between 500 and 1000 milliseconds.[1] The test word was presented on the screen for 2 seconds following the fixation, which was followed by a 200 millisecond blank screen.[2] After the presentation of the test word, participants were asked to make an old/new judgment, provide a confidence rating (1: not confident, 2: somewhat confident, 3: confident, 4: very confident), and finally choose whether they *know* or *remember* the presented word by using keyboard presses in a self-paced manner. The designated keys for the respective responses were randomized and counterbalanced for each participant (Figure 2.2).

[1] The reason for using the jittering fixation instead of a stable fixation was due to the original planning of the experiment as a behavioral pre-study for a later EEG experiment. However, the EEG stage of the study was cancelled due to postponements related to the pandemic conditions, and the study was further expanded as a fully behavioral experiment.

[2] The durations of the test stimulus and the post-blank screen were similarly set to account for the corresponding brain signals that would be acquired during the retrieval phase.

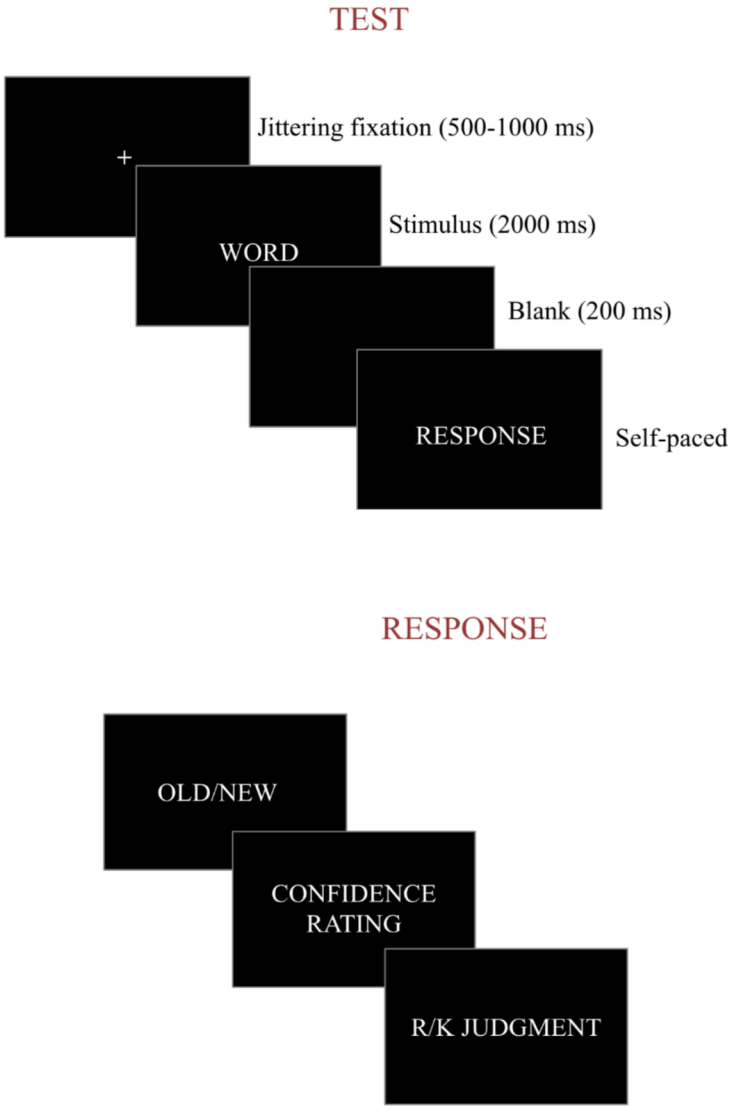

Figure 2.2 An illustration of the test phase, and the serialization of the responses.

During the response phase, confidence ratings were intentionally asked before the R-K judgments in order to make certain that the ratings reflected the confidence in the old/new judgments of the given word, rather than the confidence in the R-K judgments (Figure 2.2). Before beginning the experiment, participants were asked to put their right and left index fingers on the predetermined keys for old/new responses or R-K

responses. There was no upper limit for participants to provide a response, all test trials were self-paced.

2.2. Results

2.2.1. Data Transformation

Sensitivity (*d'*) values, criterion (*C*) values, HRs and FARs were calculated for each condition prior to the following statistical analyses (see Table 2.2 for descriptives). HRs have been calculated by dividing the total number of hits to the total number of targets, and FARs have been calculated by dividing the total number of false alarms to the total number of foils. The response bias measure *C* and the sensitivity measure *d'* were calculated by using hit and false alarm rates for each condition as explained in the Introduction chapter.

Table 2.2
d' Values in Experiment 1

	Criterion	Base Rate	*d'*
N	conservative	BRM	9
		No BRM	9
	liberal	BRM	9
		No BRM	9
Mean	conservative	BRM	1.43
		No BRM	1.44
	liberal	BRM	1.66
		No BRM	1.72
SD	conservative	BRM	0.374
		No BRM	0.68
	liberal	BRM	0.66
		No BRM	0.571

2.2.2. Between Subjects Effects

A two-way ANOVA was performed to analyze the effect of base rate and criterion on the following variables using the Jamovi software (R Core Team 2018; The jamovi project, 2019). Since the sample size was small, determining the distributions of the following variables was important for choosing an appropriate statistical method. Hence, the assumptions of normality and homogeneity of variances were controlled by performing the Shapiro-Wilk and Levene's tests.

2.2.2.1.Sensitivity

The sensitivity scores of *d'* were distributed normally according to the results of the Shapiro-Wilk test ($W = 0.97$, $p = 0.59$). Levene's test indicated equal variances (F (1, 32) = 0.65, $p = 0.59$). A two-way ANOVA was performed on *d'* values with base rate and criterion as fixed factors. The main effect of base rate on sensitivity was nonsignificant (F (1, 32) = 0.030, $MSE = 0.010$, $p = 0.864$, $\eta 2p = 0.001$), so was the main effect of criterion on sensitivity (F (1, 32) = 1.726, $MSE = 0.588$, $p = 0.198$, $\eta 2p = 0.051$). The interaction between base rate and criterion was also nonsignificant (F (1, 32) = 0.020, $MSE = 0.007$, $p = 0.889$, $\eta 2p = 0.001$).

2.2.2.2. Response Bias

The results of the Shapiro-Wilk test did not show any evidence of non-normality for response bias (W = 0.97, $p = 0.36$). Levene's test indicated equal variances (F (1, 32) = 0.19, $p = 0.91$). A two-way ANOVA was performed on *C* values with base rate and criterion as fixed factors. The main effect of criterion on response bias was significant (F (1, 32) = 6.839, $MSE = 0.477$, $p < 0.05$, $\eta 2p = 0.176$). The conservative criterion condition resulted in a larger *C* value ($M = 0.087$, $SE = 0.062$) than the liberal criterion condition ($M = -0.143$, $SE = 0.062$), as expected (Table 2.3). The main effect of base

rate on response bias was nonsignificant (F (1, 32) = 0.174, MSE = 0.012, p = 0.680, $\eta 2p$ = 0.005). The interaction between base rate and criterion was also nonsignificant (F (1, 32) = 0.260, MSE = 0.018, p = 0.613, $\eta 2p$ = 0.008).

Table 2.3
C Values in Experiment 1

	Criterion	Base Rate	*d'*
N	conservative	BRM	9
		No BRM	9
	liberal	BRM	9
		No BRM	9
Mean	conservative	BRM	0.0833
		No BRM	0.0915
	liberal	BRM	-0.102
		No BRM	-0.183
SD	conservative	BRM	0.3
		No BRM	0.274
	liberal	BRM	0.24
		No BRM	0.237

2.2.2.3.Hit and False Alarm Rates

The two-way ANOVA on the effect of base rate and criterion on the corresponding hit rates was not performed due to unequal variances between the groups as shown by the result of the Levene's test (F (1, 32) = 3.57, $p < 0.05$). For the false alarm rates, however, the Shapiro-Wilk (W = 0.966, p = 0.328) and Levene's tests (F (1, 32) = 2.05, p = 0.127) were nonsignificant. The main effects of criterion (F (1, 32) = 0.890, MSE = 0.010, p = 0.352, $\eta 2p$ = 0.027) and base rate (F (1, 32) = 0.055, MSE = 0.001, p =

0.815, $\eta 2p$ = 0.002) were found to be nonsignificant. The interaction was also nonsignificant (F (1, 32) = 0.258, MSE = 0.003, p = 0.615, $\eta 2p$ = 0.008). Hence, neither the criterion nor the base rate conditions had an effect on the resulting false alarm rates.

2.2.2.4.Confidence

The Levene's test (F (3, 32) = 0.152, p = 0.927) and the Shapiro-Wilk test (W = 0.962, p = 0.251) returned nonsignificant results, indicating equal variances and normal distribution between the groups. The results of the two-way ANOVA showed that there were no main effects of criterion (F (1, 32) = 0.170, MSE = 0.039, p = 0.683, $\eta 2p$ = 0.005) or base rate (F (1, 32) = 0.595, MSE = 0.136, p = 0.446, $\eta 2p$ = 0.018) on participants' confidence ratings. Moreover, the interaction between criterion and base rate was also nonsignificant (F (1, 32) = 0.003, MSE = 0.001, p = 0.958, $\eta 2p$ = 0.001).

2.2.3. Within Subjects Effects

For the analyses of the within subjects effects, it was necessary to calculate the respective R and K values for hit rates, false alarm rates, sensitivity and criterion variables. Since R and K responses are given only for the items that are grouped as old, there are no R and K responses associated with correct rejections or misses based on new items. This suggests that during the calculation of hit and false alarm rates, there is no way of using the number of correct rejections and misses from R and K responses. Hence, any calculation of hit rates and false alarm rates associated with R and K responses is inevitably dependent on the total number of foils, which are in no way associated with these responses. Although this leads to a lower proportion of hit and false alarms when the HR and FARs are calculated, the fact that the two responses are still bound to the same pool of foils ensure that the corresponding HRs and FARs are comparable to one another. The below formulas were used for the calculation of

HRs and FARs for R and K responses, as well as the calculation of *d'* and *C*, which inherently involve HRs and FARs in their calculations.

(3) HR (R) = Number of R Hits / (Number of R Hits + Total Number of Misses)

(4) HR (K) = Number of K Hits / (Number of K Hits + Total Number of Misses)

(5) FAR (R) = Number of R False Alarms / (Number of R False Alarms + Total Number of Correct Rejections)

(6) FAR (K) = Number of K False Alarms / (Number of K False Alarms + Total Number of Correct Rejections)

(7) d' (R) = z (HR[r]) - z (FAR[r])

(8) d' (K) = z (HR[k]) - z (FAR[k])

(9) C (R) = -0.5 * [z (HR[r]) + z (FAR[r])]

(10) C (K) = -0.5 * [z (HR[k]) + z (FAR[k])]

Note that during the calculation of HRs and FARs for R responses, some participants' false alarms could not be transformed into z-scores because they had no false alarms. In order to fix this issue, the values of 0 were replaced with 0.5/*n* where *n* is the number of noise trials (Macmillan & Kaplan, 1985).[3]

[3] This approach was suggested by Macmillan and Kaplan (1985) as a solution to the problem of extreme values (i.e. 0 and 1) when calculating z-scores for hit rates and false alarms. The solution involves adjusting these extreme rates by 0.5 / n when the rate is 0, and by (n - 0.5) / n, when the rate is 1. The n refers to the number of signal or noise trials depending on whether the adjusted rate refers to a hit rate or a false alarm rate, respectively.

2.2.3.1. Number of R and K Responses

A mixed ANOVA was performed to compare the number of R and K responses as a function of criterion and base rate. The sphericity and homogeneity assumptions were met based on the results of the test of sphericity and Levene's test (number of R responses: F (3, 32) = 1.73, p = 0.177; number of K responses: F (3, 32) = 1.05, p = 0.385). The results showed a significant effect of criterion on R and K responses (F (1, 32) = 34.667, MSE = 14706.13, p < 0.001, $\eta 2p$ = 0.520), and a significant criterion by base rate interaction (F (1, 32) = 6.719, MSE = 2850.13, p < 0.05, $\eta 2p$ = 0.174). The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that the mean difference between R and K responses were significant between the conservative and liberal conditions only when there was a base rate manipulation (t (32) = 6.00, MD = 41.2, SE = 6.87, p < 0.001; see Figure 2.3).

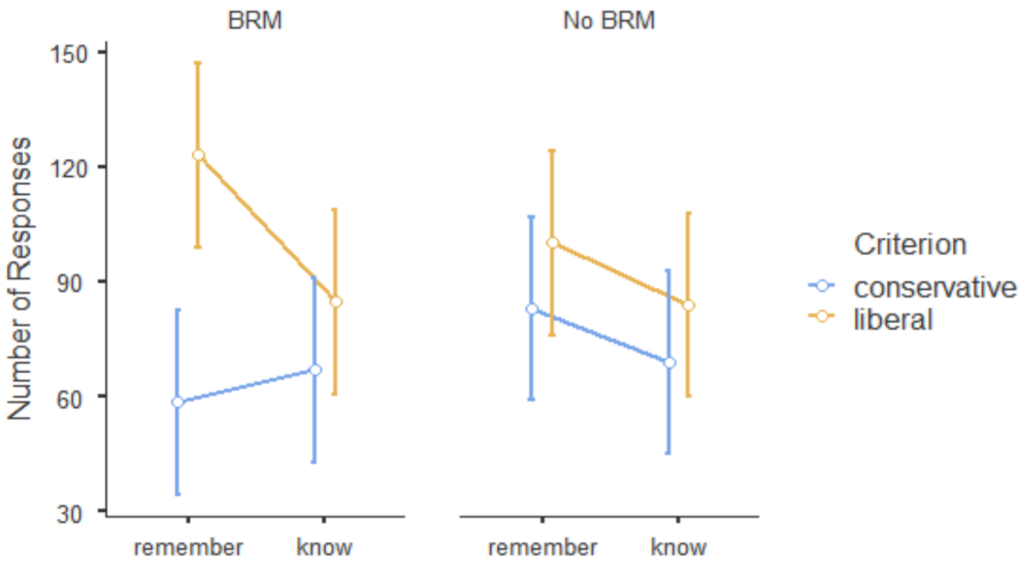

Figure 2.3 Number of R and K responses in Experiment 1.

2.2.3.2. Proportion of R and K Responses

A mixed ANOVA was performed to compare the proportion of R and K responses as a function of criterion and base rate. The sphericity and homogeneity assumptions were met based on the results of the test of sphericity and Levene's test (proportion of R responses: F (3, 32) = 1.08, p = 0.373; number of K responses: F (3, 32) = 1.08, p = 0.373). The results revealed a nonsignificant effect of response type (F (1, 32) = 1.037, MSE = 0.07, p = 0.316, $\eta 2p$ = 0.031), and nonsignificant interactions between response type and criterion (F (1, 32) = 0.962, MSE = 0.07, p = 0.334, $\eta 2p$ = 0.029), response type and base rate (F (1, 32) = 0.0595, MSE = 0.004, p = 0.316, $\eta 2p$ = 0.002), and response type, criterion and base rate (F (1, 32) = 0.604, MSE = 0.04, p = 0.443, $\eta 2p$ = 0.019). The proportion of R and K responses for each condition is illustrated in Figure 2.4.

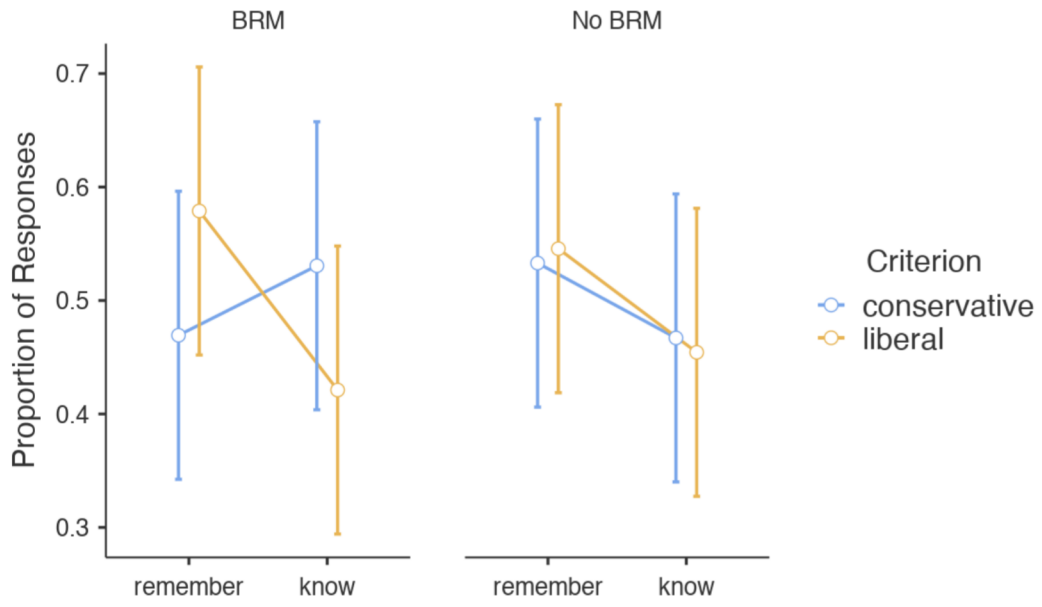

Figure 2.4 Proportion of R and K responses in Experiment 1.

2.2.3.3. Confidence in R and K responses

In order to compare the confidence levels of participants' "remember" and "know" responses, a mixed ANOVA was performed to examine confidence levels of familiarity and recollection as a function of criterion and base rate. Accordingly, respective confidence levels of "remember" and "know" responses were the repeated measures factors, and the criterion and base rate conditions were the between subjects factors. The sphericity and homogeneity assumptions were met based on the results of the test of sphericity and Levene's test (remember confidence: F (3, 31) = 1.858, p = 0.157; know confidence: F (3, 31) = 0.605, p = 0.617). The results showed a significant change between the confidence levels of R and K responses (F (1, 31) = 54.860, MSE = 20.477, p < 0.001, $\eta 2p$ = 0.639), according to which R responses were given with higher confidence than K responses. However, the confidence by criterion interaction (F (1, 31) = 0.046, MSE = 0.017, p = 0.832, $\eta 2p$ = 0.001), the confidence by base rate interaction (F (1, 31) = 0.214, MSE = 0.080, p = 0.647, $\eta 2p$ = 0.007) and the confidence, criterion, and base rate interaction were found to be nonsignificant (F (1, 31) = 2.758, MSE = 1.029, p = 0.107, $\eta 2p$ = 0.082) suggesting that the found difference in confidence scores is independent of response bias and base rate manipulations (Figure 2.5).

The proportion of confidence ratings showed an ascending trend with response frequencies increasing as confidence increases (Figure 2.6). When the responses were divided in to R, K and new responses, the frequency of lower ratings was higher for K responses than it was for R responses whereas the frequency of a rating of 4 was higher for the R responses than it was for the K responses. Moreover, the frequency of new responses was at a minimum when the confidence was at its highest (Figure 2.7).

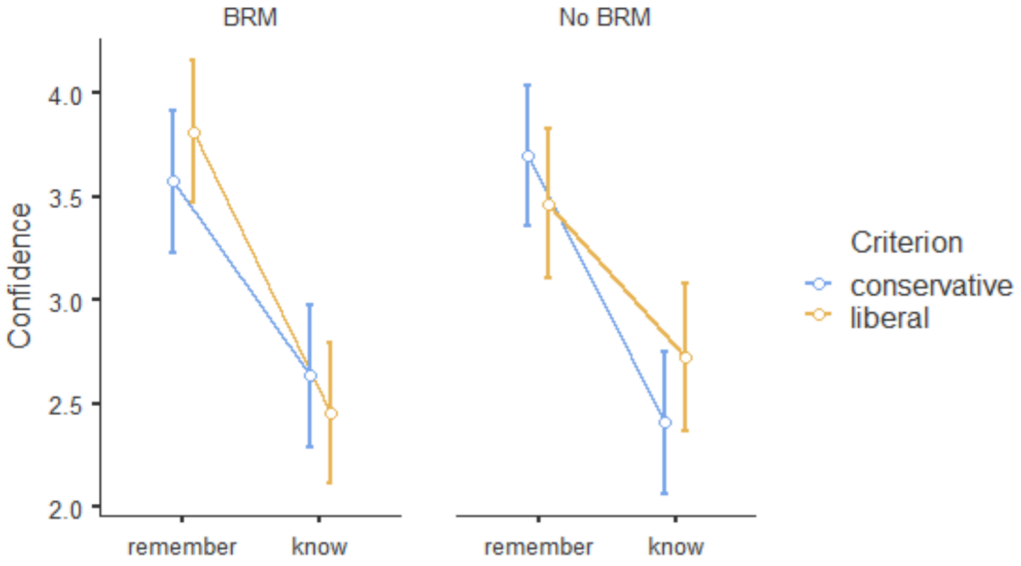

Figure 2.5 Confidence levels of R and K responses, grouped by conservative and liberal response criteria.

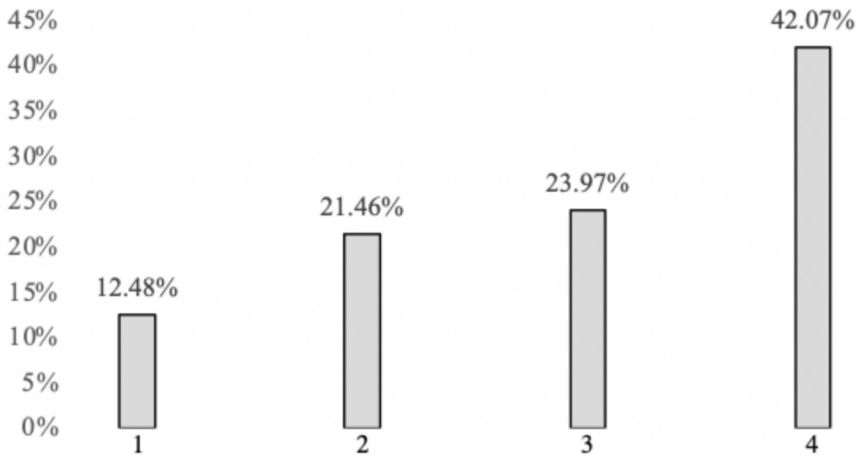

Figure 2.6 Percentage of confidence ratings as a function of confidence.

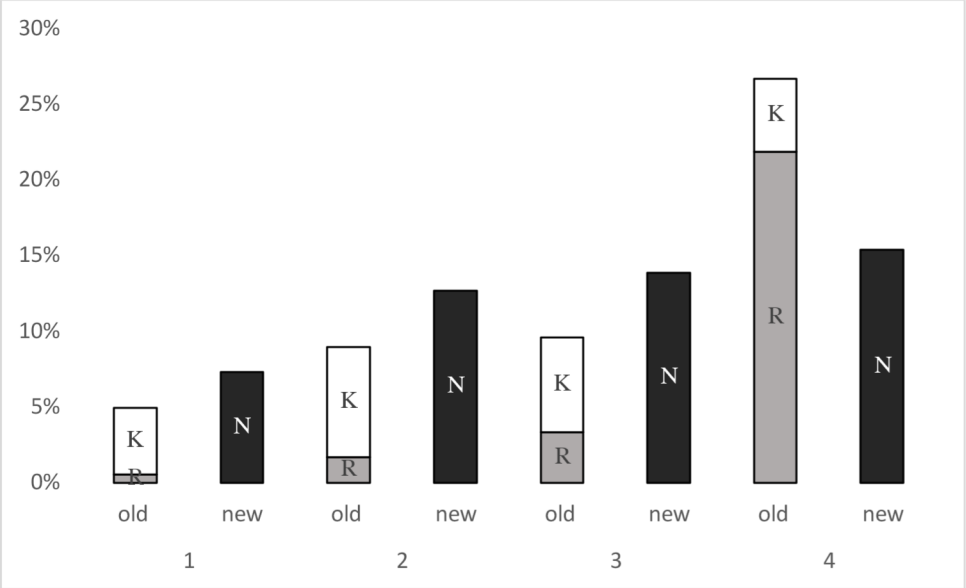

Figure 2.7 Percentage of confidence ratings as a function of remember (R), know (K), or new (N) responses.

2.2.3.4. *d'* and *C* Values for R and K Responses

A repeated measures ANOVA was performed on the sensitivity (*d'*) scores of R and K responses. The sensitivity scores for the R and K responses were calculated based on the HR and FAR calculations of the two response types, as explained above. The sphericity and homogeneity assumptions were met (Levene's F (3, 32) = 0.936, p = 0.435 in *d'* for R; F (3, 32) = 0.789, p = 0.509 in *d'* for K). The results revealed a significant difference between the *d'* scores (F (1, 35) = 38.3, MSE = 19.417, $p < 0.001$, $\eta 2p$ = 0.522) where R responses (M = 2.12, SE = 0.124, *95% CI* = [1.873, 2.37]) had greater *d'* than K responses (M = 1.08, SE = 0.124, *95% CI* = [0.834, 1.33]).

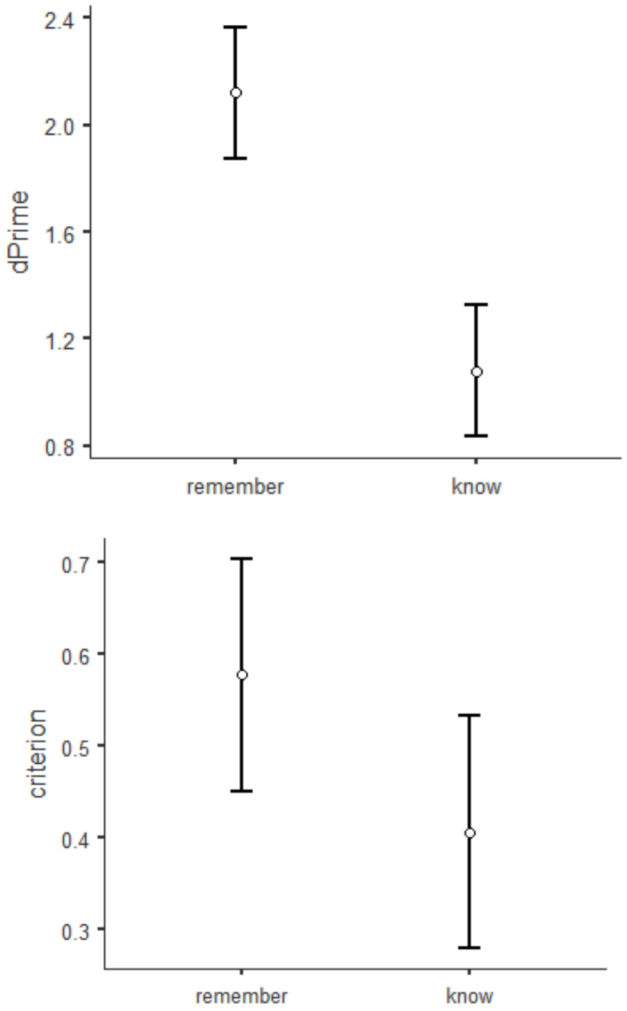

Figure 2.8 Mean sensitivity (*d'*) scores, and mean criterion (*C*) values for "remember" and "know" responses.

As for the criterion values of the R and K responses, HR and FARs for the respective response types were utilized for the calculation of the *C* scores, similar to the calculation of *d'*. The results of the repeated measures ANOVA revealed a significant difference between the *C* values of R and K responses (F (1, 35) = 4.66, MSE = 0.527, $p < 0.05$, $\eta 2p$ = 0.117) where R responses (M = 0.577, SE = 0.064, *95% CI* = [0.450, 0.704]) were based on a more conservative criterion than K responses (M = 0.406, SE = 0.064, *95% CI* = [0.279, 0.533]).

2.2.3.5.Hits and False Alarms for K Responses

In order to compare the mean hit and false alarm values of the K responses as a function of criterion and base rate, a mixed ANOVA was performed on the corresponding number of hits and false alarms for the K responses. The K hits and K false alarms were set as the repeated measures factor, and the criterion and base rate variables were set as between subject factors. The sphericity assumption was met due to the repeated measure having only two levels (hits and false alarms). Levene's test of homogeneity of variances was met for the K hits (F (3, 32) = 2.25, p = 0.101), but not for K false alarms (F (3, 32) = 2.97, p = 0.047). However, considering the small sample size of the data set as well as the data being raw, the mixed ANOVA was still performed given the robustness of the test.

Table 2.4

Number of K hits and false alarms in Experiment 1

					95% Confidence Interval	
Base Rate	Criterion	KNOW	Mean	SE	Lower	Upper
BRM	conservative	HIT	27.6	7.23	13.11	42.0
		FALSE ALARM	39.3	7.23	24.89	53.8
	liberal	HIT	63.4	7.23	49.00	77.9
		FALSE ALARM	21.2	7.23	6.78	35.7
No BRM	conservative	HIT	39.4	7.23	25.00	53.9
		FALSE ALARM	29.7	7.23	15.22	44.1
	liberal	HIT	55.6	7.23	41.11	70.0
		FALSE ALARM	28.2	7.23	13.78	42.7

According to the results of the ANOVA, there was a significant difference between K hits and K false alarms (F (1, 32) = 12.240, MSE = 5134.2, $p < 0.01$, $\eta 2p$ = 0.280).

Moreover, there was a significant interaction between K responses of hits and false alarms, and the criterion (F (1, 32) = 13.934, MSE = 5760.2, p < 0.001, $\eta 2p$ = 0.303; see Figure 2.9). The interaction between the K responses and base rate was nonsignificant (F (1, 32) = 0.121, MSE = 50.0, p = 0.730 $\eta 2p$ = 0.004), so was the interaction between K responses, criterion, and base rate (F (1, 32) = 3.615, MSE = 1494.2, p = 0.066, $\eta 2p$ = 0.101). The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that there were more hits in the liberal condition than there were in the conservative condition (t (32) = 3.97, MD = 26.0, SE = 7.23, p < 0.01). Moreover, the number of hits were greater than the number of false alarms in the liberal condition (t (32) = 5.132, MD = 34.778, SE = 6.78, p < 0.001). Finally, the hits in the liberal condition were greater than the false alarms in the conservative condition (t (32) = 3.459, MD = 25.0, SE = 7.23, p < 0.01).

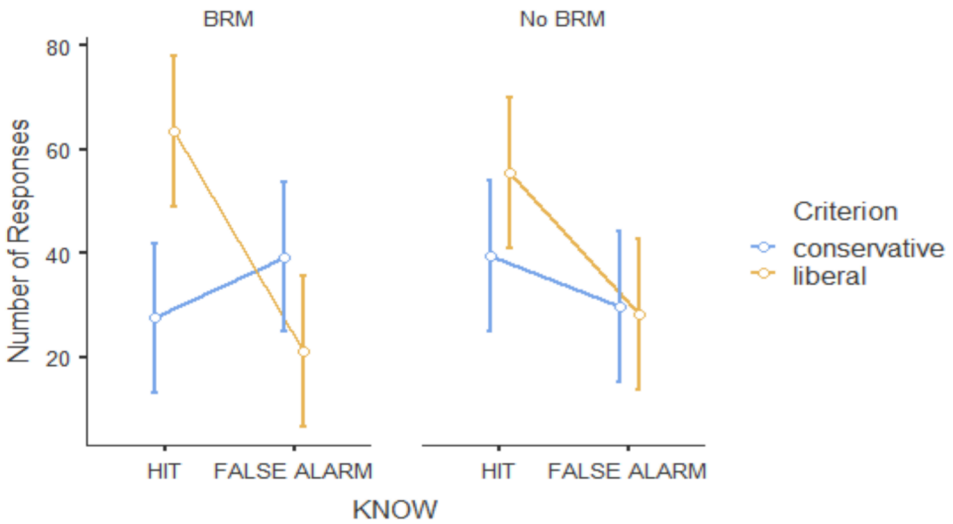

Figure 2.9 Number of "know" hits and false alarms, grouped by conservative and liberal response criteria, and base rate condition.

2.2.3.6. Hits and False Alarms for R Responses

In order to compare the mean hit and false alarm values of the R responses as a function of criterion and base rate, a mixed ANOVA was performed on the corresponding number of hits and false alarms for the R responses. The R hits and R false alarms were set as the repeated measures factor, and the criterion and base rate variables were set as between subject factors. The sphericity assumption was met due to the repeated measure having only two levels (hits and false alarms). Levene's test of homogeneity of variances was met for the R hits (F (3, 32) = 2.15, p = 0.113), but not for R false alarms (F (3, 32) = 5.97, p = 0.002). However, considering the small sample size of the data set as well as the data being raw, the mixed ANOVA was still performed given the robustness of the test.

Table 2.5
Number of R hits and false alarms in Experiment 1

					95% Confidence Interval	
Base Rate	Criterion	REMEMBER	Mean	SE	Lower	Upper
BRM	conservative	Hit	46.00	9.25	27.52	64.5
		False Alarm	12.44	9.25	−6.04	30.9
	liberal	Hit	119.78	9.25	101.29	138.3
		False Alarm	3.22	9.25	−15.26	21.7
No BRM	conservative	Hit	77.11	9.25	58.63	95.6
		False Alarm	6.00	9.25	−12.48	24.5
	liberal	Hit	83.89	9.25	65.40	102.4
		False Alarm	16.22	9.25	−2.26	34.7

According to the results of the ANOVA, there was a significant difference between R hits and R false alarms (F (1, 32) = 117.758, MSE = 93889, $p < 0.001$, $\eta 2p = 0.786$). Moreover, there was a significant interaction between R responses of hits and false alarms, and the criterion (F (1, 32) = 8.930, MSE = 7120, $p < 0.01$, $\eta 2p = 0.218$). The interaction between the R responses and base rate was nonsignificant (F (1, 32) = 0.181, MSE = 144, p = 0.673 $\eta 2p$ = 0.006). However, the interaction between R responses, criterion, and base rate was significant (F (1, 32) = 10.544, MSE = 8407, $p < 0.01$, $\eta 2p = 0.248$). The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that the number of hits were greater than the number of false alarms when the criterion was liberal, regardless of the base rate condition (t (32) = 8.756, MD = 116.56, SE = 13.3, p < 0.001 when the base rate is manipulated; t (32) = 5.084, MD = 67.67, SE = 13.3, p < 0.001 when there is no base rate manipulation).

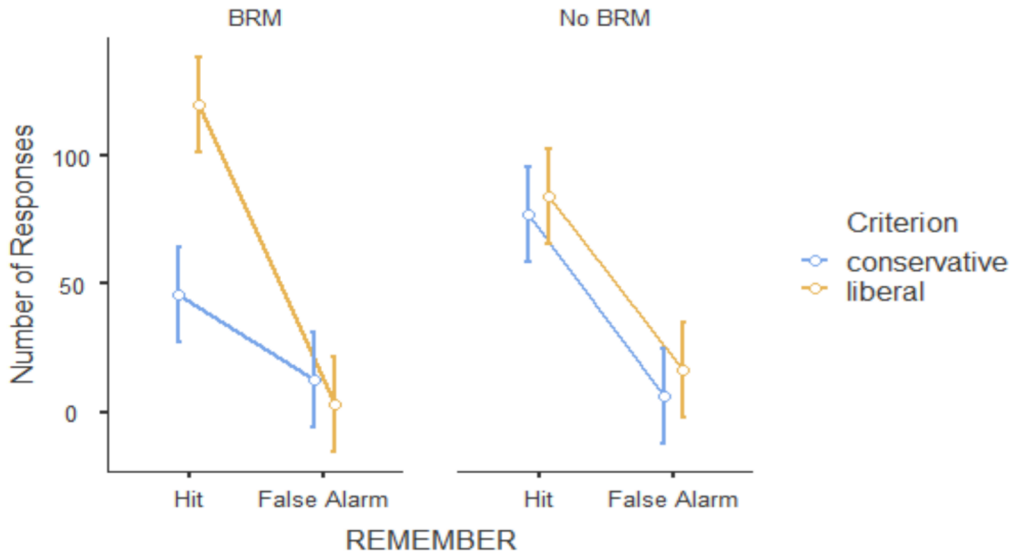

Figure 2.10 Number of "know" hits and false alarms, grouped by conservative and liberal response criteria, and base rate condition.

The number of hits were greater than the number of false alarms in the conservative condition when there was no base rate manipulation (t (32) = 5.342, MD = 71.11, SE = 13.3, p < 0.001), but it was nonsignificant under the base rate condition (t (32) = 2.521, MD = 33.56, SE = 13.3, p = 0.473). Moreover, the number of hits were greater in the liberal condition than they were in the conservative condition when there was a base rate manipulation (t (32) = 5.638, MD = 73.78, SE = 13.1, p < 0.001); however, the same effect was not significant when there was no base rate manipulation (t (32) = 0.518, MD = 6.78, SE = 13.1, p = 1).

2.2.3.7.Hit Rates and False Alarm Rates for K Responses

In order to compare the mean HRs and FARs of the K responses as a function of criterion and base rate, a mixed ANOVA was performed on the corresponding HRs and FARs for the K responses. The "know" HRs and FARs were set as the repeated measures factor, and the criterion and base rate variables were set as between subject factors. The sphericity assumption was met due to the repeated measure having only two levels (hit rates and false alarm rates). Levene's test of homogeneity of variances was met for both K-HRs (F (3, 32) = 0.50, p = 0.685), and K-FARs (F (3, 32) = 1.788, p = 0.169).

The results of the mixed ANOVA showed a significant difference between K-HRs and K-FARs (F (1, 32) = 156.876, MSE = 2.249, p < 0.001, $\eta 2p$ = 0.831). Moreover, there was a significant interaction between K responses of HRs and FARs, and the criterion (F (1, 32) = 5.205, MSE = 0.075, p < 0.05, $\eta 2p$ = 0.140; see Figure 2.11). The interaction between the K responses and base rate was nonsignificant (F (1, 32) = 0.460, MSE = 0.007, p = 0.503 $\eta 2p$ = 0.014), so was the interaction between K responses, criterion, and base rate (F (1, 32) = 1.285, MSE = 0.018, p = 0.265, $\eta 2p$ = 0.039). There was also a main effect of criterion on the corresponding HRs and FARs (F (1, 32) = 6.666, MSE = 0.126, p < 0.05, $\eta 2p$ = 0.172). The post-hoc tests were carried out using Bonferroni adjustment for p values.

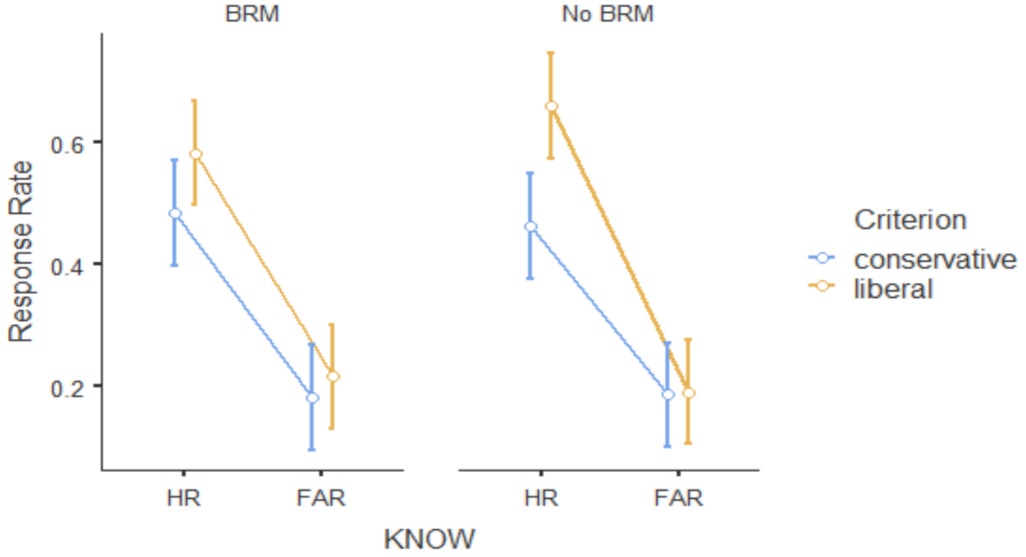

Figure 2.11 HRs and FARs for K responses, grouped by conservative and liberal response criteria, and base rate condition.

Table 2.6

HRs and FARs for K Responses in Experiment 1

					95% Confidence Interval	
Base Rate	Criterion	KNOW	Mean	SE	Lower	Upper
BRM	conservative	HIT	27.6	7.23	13.11	42.0
		FALSE ALARM	39.3	7.23	24.89	53.8
	liberal	HIT	63.4	7.23	49.00	77.9
		FALSE ALARM	21.2	7.23	6.78	35.7
No BRM	conservative	HIT	39.4	7.23	25.00	53.9
		FALSE ALARM	29.7	7.23	15.22	44.1
	liberal	HIT	55.6	7.23	41.11	70.0
		FALSE ALARM	28.2	7.23	13.78	42.7

The results of the post-hoc comparisons revealed that the HR was greater than the FAR when the criterion was liberal, regardless of the base rate condition (t (32) = 2.58, MD = 0.08, SE = 0.032, $p < 0.05$). Moreover, the overall K-HR was greater than the K-FAR in the conservative condition (t (32) = 12.5, MD = 0.35, SE = 0.028, $p < 0.001$). The K-HR was greater than the K-FAR under both conservative (t (32) = 7.243, MD = 0.29, SE = 0.04, $p < 0.001$) and liberal conditions (t (32) = 10.47, MD = 0.42, SE = 0.04, $p < 0.001$). Finally, the overall HR was greater in the liberal condition (t (32) = 3.45, MD = 0.15, SE = 0.04, $p < 0.01$).

2.2.3.8.Hit Rates and False Alarm Rates for R Responses

In order to compare the mean HRs and FARs of R responses as a function of criterion and base rate, a mixed ANOVA was performed on the corresponding number of HRs and FARs for the R responses. The R-HRs and R-FARs were set as the repeated measures factor, and the criterion and base rate variables were set as between subject factors. The sphericity assumption was met due to the repeated measure having only two levels (HRs and FARs). Levene's test of homogeneity of variances was met for R-HRs (F (3, 32) = 2.49, p = 0.078), but not for R-FARs (F (3, 32) = 3.82, p = 0.019). However, considering the small sample size of the data set as well as the data being raw, the mixed ANOVA was still performed given the robustness of the test.

According to the results of the ANOVA, there was a significant difference between R-HRs and R-FARs (F (1, 32) = 291.016, MSE = 6.346, $p < 0.001$, $\eta 2p$ = 0.901). The interaction between R responses of HRs and FARs, and the criterion was nonsignificant (F (1, 32) = 2.223, MSE = 0.048, p = 0.146, $\eta 2p$ = 0.065), so was the interaction between R response proportions and base rate (F (1, 32) = 0.045, MSE = 0.001, p = 0.833 $\eta 2p$ = 0.001), and the interaction between R response proportions, criterion, and base rate (F (1, 32) = 0.098, MSE = 0.002, p = 0.757, $\eta 2p$ = 0.003; see Figure 2.12).

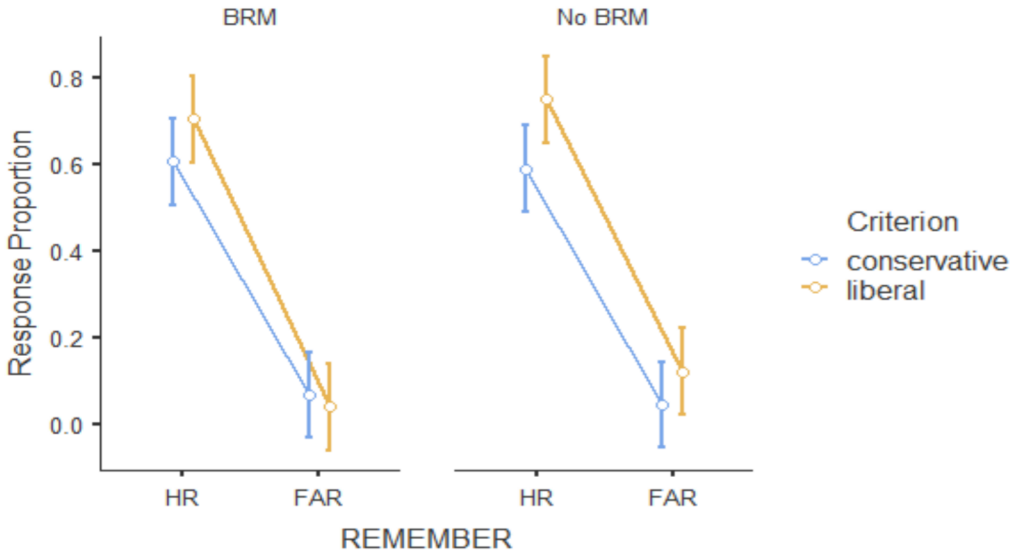

Figure 2.12 HRs and FARs for R responses grouped by conservative and liberal response criteria, and base rate condition.

Finally, the effect of criterion on the R-HRs and R-FARs was significant (F (1, 32) = 4.635, MSE = 0.105, $p < 0.05$, $\eta 2p$ = 0.127) in favor of HRs. The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that HRs were greater than FARs under both criterion conditions, regardless of base rate (t (32) = 17.1, MD = 0.59, SE = 0.03, p < 0.001). Moreover, overall HRs and FARs were greater in the liberal condition (t (32) = 2.15, MD = 0.08, SE = 0.04, p < 0.05; see Table 2.7).

Table 2.7

HRs and FARs for R Responses in Experiment 1

					95% Confidence Interval	
Base Rate	Criterion	REMEMBER	Mean	SE	Lower	Upper
BRM	conservative	HR	0.6056	0.0497	0.5064	0.705
		FAR	0.0672	0.0497	−0.0320	0.166
	liberal	HR	0.7031	0.0497	0.6039	0.802
		FAR	0.0392	0.0497	−0.0600	0.138
No BRM	conservative	HR	0.5897	0.0497	0.4905	0.689
		FAR	0.0443	0.0497	−0.0549	0.144
	liberal	HR	0.7485	0.0497	0.6493	0.848
		FAR	0.1211	0.0497	0.0219	0.220

2.3. Discussion

The findings of Experiment 1 were useful to see whether the experimental manipulations were working, and to understand how criterion shift could be performed under optimal conditions in a between-subjects design. First, the results of Experiment 1 found no difference in the sensitivity scores between any conditions. This means that the found differences in R and K responses in terms of criterion and base rate manipulations cannot be attributed to participants' performance, that is, the ability to discriminate between targets and lures. The sensitivity of R responses, however, was greater than the sensitivity for K responses. Second, the criterion value between the conservative and liberal conditions were found to be significantly different, suggesting that the applied criterion manipulation worked independently of whether there was a base rate manipulation. However, in terms of the number of responses given by participants, there was a difference between the number of R and K responses only when there was a base rate manipulation. In this regard, there were more R responses when the criterion was liberal, and more K responses when the criterion was conservative only for participants whose test lists had an actual criterion manipulation.

Note that the confidence levels of the R responses were greater than the confidence levels of the K responses under every condition. Therefore, considering the differing number of R and K responses under different criterion conditions, the findings can be interpreted in terms of a confidence argument. Participants were more likely to respond confidently when the criterion was liberal, because their chances of hitting a target was greater, which translated into a greater number of R responses that were significantly more high-confidence than K responses. Although it can be argued that such interpretation contradicts with the finding that confidence did not differ between the liberal and conservative responses (when the number of R and K responses was held constant), a counterargument to this can be seen in the mean confidence scores of the two response types. Despite the lower confidence scores for the K responses, the mean confidence score for R responses was 3.68, which indicates a ceiling effect. Hence, despite the lower number of R responses in the conservative condition, the extremely high R scores may have pulled the mean higher, resulting in a similar overall confidence between liberal and conservative conditions. However, no significant difference was found between the corresponding proportions of R and K responses under any conditions, which creates drawbacks in interpreting the results in the abovementioned direction. More specifically, given the fact that proportions yield more valid measures for comparison than frequencies, and that the proportion of responses did not differ from one another, suggest that R and K response tendencies may not be affected by a base rate manipulation or response bias.

In terms of the hits and false alarms for K responses, there were more hits than false alarms when the criterion was liberal. Moreover, hits were greater in the liberal condition than they were in the conservative condition. The greater number of hits when the criterion is more lenient is expected with such manipulations. As for the greater number of hits as compared to false alarms, Donaldson's model argues for the opposite results when the criterion is liberal. This is because there is a smaller area for hits than false alarms, when the criterion is liberal according to the two-threshold model. The present experiment yielded results that are not supporting this model. When the hit and false alarm rates were calculated for the K responses, however, the

trend was in the direction of greater HRs than FARs for both liberal and conservative conditions. This clear direction of these results may be suggesting drawbacks with the calculation of the proportion of responses. There may be several explanations for the contrasting evidence in terms of the number of hits and false alarms for K responses, and the consistent trend of higher proportion of hit rates, which will be further discussed in the General Discussion.

For the hit and false alarms of R responses, hits were greater in the liberal condition than they were in the conservative condition only when there was a base rate manipulation. The greater number of hits in the liberal condition is one of the main indicators of a leaner criterion, and the lack of a difference between hits when there is no base rate manipulation suggests that the criterion shift might not have occurred under those circumstances. It was found that hits were greater than false alarms when the criterion was liberal, similar to the results of the K responses. Moreover, hits were greater than false alarms in the conservative condition only when there was no base rate manipulation. The same effect was not present when there was a base rate manipulation. Note that when the hit and false alarm rates were calculated, there was the same trend as in K responses, which is characterized by a greater number of hits than false alarms for both conservative and liberal conditions.

The transformation of the number of hits and false alarms led to distortions in the actual proportion of the responses due to the different number of targets and foils in each condition. For example, when the false alarm rate was calculated for the conservative criterion, the denominator was larger for responses in the conservative condition since there were more lures than targets. Contrariwise, when the false alarm rate was calculated for the liberal criterion, the denominator was smaller. This resulted in an inflated hit rate in the liberal condition as compared to the conservative condition, which may explain the unidirectional trend in the hit and false alarm rate analyses. However, the same hit and false alarm rates were used in the calculation of *C* values as well. The fact that a distinction between the R and K responses is still found in terms

of differences in *C* values, despite the drawbacks of this calculation, further supports the idea of a high criterion R threshold and a low criterion K threshold.

The results of Experiment 1 suggested that the criterion manipulation is in fact presenting the intended shift in the *C* values. However, some of the fundamental changes expected with the criterion shift were not observed under some of the conditions when there was no base rate manipulation such as the lack of a difference between the hit rates of liberal and conservative conditions, and the lack of a crossover effect on the number of R and K responses in conservative and liberal conditions. Considering this, along with the constraint that an additional variable will create on the sample size, it has been decided not to include the base rate condition in Experiment 2. This decision was further supported given that this decision would not affect the criterion effect (i.e. the desired *C* values are observed with or without the base rate manipulation), which is the main point of the experiments, but that it would allow the familiarity and recollection processes to be examined in a more simplified experimental design with larger samples in each condition.

CHAPTER 3

EXPERIMENT 2

3.1. Method

Experiment 2 was based on the same rationale as Experiment 1 except the fact that minor methodological changes have been made in Experiment 2. First, the base rate manipulation condition has been removed from Experiment 2 given the results of Experiment 1, in which most of the criterion effects were observed when the base rate was manipulated (e.g. number of R and K responses, lack of a difference between the hits among the criterion conditions). Given this, the no base rate manipulation condition has been removed in order to prevent further grouping of the participants, and to have a larger sample size in each condition. Second, in order to improve overall performance, and to ensure that the criterion manipulation has the intended effect, feedback was given after each test item.

3.1.1. Participants

The initial goal for the number of participants in Experiment 2 was 150, which was driven by the calculation of the sample size through a power analysis based on previous studies. However, due to difficulties in finding participants under the pandemic conditions, a total number of 62 undergraduate students could be recruited from Middle East Technical University (METU) to participate in the study. Students received a partial course credit in return for their participation. All participants

provided written consent prior to the beginning of the experiment. The data from two of the participants were excluded from the final analysis. One participant's performance was below the chance level (i.e. 50%), and three standard deviations lower than the mean performance among other participants. The other participant's sensitivity level, measured by *d'*, and hit rate were three standard deviations away from the mean performance in the respective categories. Therefore, after removing the data from the two participants, the complete data from 60 participants (M age = 21.8, SD age = 1.83) were used in the analysis There were 10 males and 52 females, and 7 left-handed and 55 right-handed people All participants were native Turkish speakers with normal color-vision and normal or corrected-to-normal visual acuity.

3.1.2. Materials

The same word pool as in Experiment 1 was used. The number and the proportion of words used in each block and condition were also the same as Experiment 1. Accordingly, 559 words were randomly sampled for the conservative condition, and 431 words were randomly sampled for the liberal condition in each experimental session from the Turkish Word Norms (Tekcan & Göz, 2005).

3.1.3. Design and Procedure

The procedure was the same as the procedure in Experiment 1. However, there were minor differences in the design of the experiment. First, the base rate manipulation was applied to all participants in Experiment 2, given the results of Experiment 1 that there is no criterion shift when there is no base rate manipulation. Second, feedback was given after each test item to improve performance and the effectiveness of the criterion manipulation. The rationale for the addition of a feedback was based on the results of a study by Koop, Criss and Malmberg (2015), in which participants' performance on standard (i.e. equal number of targets and foils) and pure (i.e. test lists consists only of targets or foils) test lists were identical when there was no feedback

regarding the target to foil proportion, yet an improvement in performance was observed with an accurate feedback regarding these proportions.

The instructions and the flow of the experiment was the same with Experiment 1. One of the major differences between the two experiments was that there was a two year break between the two experiments, which meant that Experiment 1 was pre-pandemic and Experiment 2 was post-pandemic. There are no preconceived predictions regarding the effect of the pandemic on the results based on the given theoretical framework. Hence, any differences that may occur between the two experiments are expected to occur due to differences in the experimental design.

3.2. Results

3.2.1. Data Transformation

Similar to Experiment 1, sensitivity (*d'*) values, criterion (*C*) values, HRs and FARs were calculated for each condition prior to the following statistical analyses. The same method of calculation was used for each value of interest.

3.2.2. Between Subjects Effects

A one-way ANOVA was performed to analyze the effect of criterion on the following variables using the Jamovi software. Assumptions of normality and homogeneity of variances were controlled by performing the Shapiro-Wilk and Levene's tests.

3.2.2.1. Sensitivity

Levene's test indicated equal variances (F (1, 58) = 1.91, p = 0.173). The sensitivity scores of *d'* were not normally distributed according to the results of the Shapiro-Wilk test (W = 0.956, p = 0.031). However, a one-way ANOVA was still performed to examine the effect of criterion manipulation on corresponding sensitivity scores. The effect of criterion on sensitivity was found to be significant (F (1, 58) = 31.8, MSE = 10.342, p < 0.001, $\eta 2p$ = 0.354). The sensitivity, measured by *d'* was higher in the conservative condition (M = 1.99, SE = 0.106, *95% CI* = [1.778, 2.20]) than it was in the liberal condition (M = 1.16, SE = 0.102, *95% CI* = [0.954, 1.36]).

3.2.2.2. Response Bias

A one-way ANOVA was performed to examine the effect of criterion manipulation on the response bias scores of participants, measured by *C*. The assumption of homogeneity of variances between the groups was met based on the result of the Levene's test (F (1, 58) = 0.532, p = 0.469). The assumption of normality was also met based on the result of Shapiro-Wilk (W = 0.972, p = 0.174). The results of the one-way ANOVA showed that there is a significant effect of criterion on the corresponding *C* values (F (1, 58) = 62.3, MSE = 3.360, p < 0.001, $\eta 2p$ = 0.518). This suggested that the intended criterion manipulation provided a shift in participants' response bias during the recognition test through a criterion shift, as hypothesized in the SDT model. Accordingly, *C* was higher in the conservative condition (M = 0.224, SE = 0.043, *95% CI* = [0.138, 0.311]), indicating a stricter criterion, as compared to the liberal condition (M = -0.249, SE = 0.042, *95% CI* = [-0.333, -0.166]).

3.2.2.3. Hit and False Alarm Rates

As in Experiment 1, the result of the Levene's test was significant (F (1, 58) = 10.8, p = 0.002), implying unequal variances among hit rates. The result of the Shapiro-Wilk test was also significant (W = 0.932, p = 0.003). For the false alarm rates, although the Shapiro-Wilk test returned a nonsignificant result (W = 0.977, p = 0.309), the result of the Levene's test was significant (F (1,58) = 23.1, p < 0.001). However, an ANOVA was still performed to examine the HRs and FARs as a function of criterion. For the HRs, the effect of criterion on the corresponding hit rates was nonsignificant (F (1, 58) = 3.13, MSE = 0.05, p = 0.082, $\eta 2p$ = 0.050). However, for the FARs, there was a significant effect of criterion on the false alarm rates (F (1, 58) = 51.0, MSE = 0.93, p < 0.001, $\eta 2p$ = 0.460). The post-hoc test was carried out using Bonferroni adjustment for p values. The result of the post-hoc comparison revealed that the FAR was higher in the liberal condition than it was in the conservative condition (t (58) = 7.14, MD = 0.245, SE = 0.03, p < 0.001).

3.2.2.4. Confidence

The Levene's test (F (1, 58) = 3.82, p = 0.056) and the Shapiro-Wilk test (W = 0.963, p = 0.068) returned nonsignificant results, indicating equal variances and normal distribution between the groups. The results of the two-way ANOVA showed no significant effect of criterion (F (1, 58) = 0.090, MSE = 0.022, p = 0.765, $\eta 2p$ = 0.002) on the corresponding confidence levels.

3.2.3. Within Subjects Effects

The same calculations of HRs, FARs, d', and C were used in the within subject analyses as in Experiment 2. For the calculation of FARs in data sets with no false

alarm rates, the same corrections were made as in Experiment 1 by replacing 0 with 0.5/n where *n* is the number of noise trials (Macmillan & Kaplan, 1985).

3.2.3.1. Number of R and K Responses

A mixed ANOVA was performed to compare the number of R and K responses as a function of criterion and base rate. The sphericity and homogeneity assumptions were partially met based on the results of the test of sphericity and Levene's test (number of R responses: F (1, 58) = 3.03, p = 0.087; number of K responses: F (1, 58) = 8.40, p = 0.005). The mixed ANOVA was still performed to examine whether the number of R and K responses differed depending on criteria. The results showed a significant effect of criterion on the R and K responses (F (1, 58) = 545, MSE = 95844, p < 0.001, $\eta 2p$ = 0.904). The post-hoc test was carried out using Bonferroni adjustment for p values. The result of the post-hoc comparison revealed that the mean difference between R and K responses were greater in the liberal condition than it was in the conservative condition (t (58) = 23.3, MD = 56.6, SE = 2.42, p < 0.001; see Table 3.1)

Table 3.1
Number of R and K Responses in Experiment 2

				95% Confidence Interval	
Criterion	Number of Responses	Mean	SE	Lower	Upper
conservative	remember	52.6	5.60	41.4	63.8
	know	50.1	5.60	38.9	61.3
liberal	remember	116.3	5.44	105.5	127.1
	know	99.5	5.44	88.7	110.4

3.2.3.2. Proportion of R and K Responses

A mixed ANOVA was performed to compare the proportion of R and K responses as a function of criterion. The sphericity and homogeneity assumptions were met based on the results of the test of sphericity and Levene's test (proportion of R responses: F (1, 60) = 3.32, p = 0.074; number of K responses: F (1, 60) = 3.32, p = 0.074). The results revealed a nonsignificant effect of response type (F (1, 60) = 1.576, MSE = 0.12, p = 0.214, $\eta 2p$ = 0.026), and a nonsignificant response type by criterion interaction (F (1, 60) = 0.144, MSE = 0.01, p = 0.706, $\eta 2p$ = 0.002). The proportion of R and K responses for each condition is illustrated in Figure 3.1

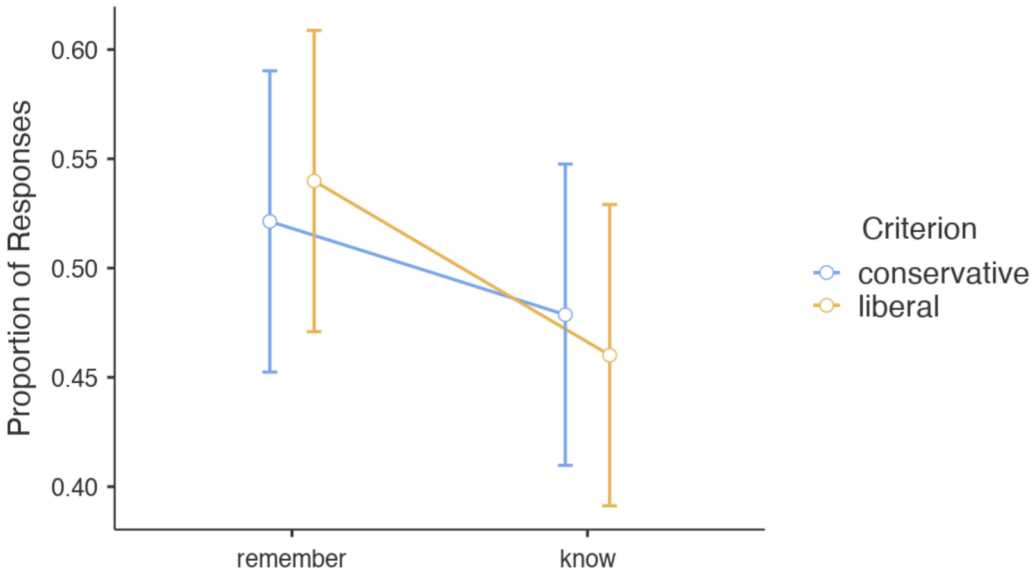

Figure 3.1 Proportion of R and K responses in Experiment 2.

3.2.3.3. Confidence in R and K responses

In order to compare the confidence levels of participants' "remember" and "know" responses, a mixed ANOVA was performed to examine confidence levels of

familiarity and recollection as a function of criterion. Accordingly, respective confidence levels of "remember" and "know" responses were the repeated measures factor, and the criterion WAS the between subjects factor. Whereas the sphericity assumption was met, the homogeneity assumptions were not met based on the results of the Levene's test (remember confidence: F (1, 58) = 35.57, p < 0.001; know confidence: F (1, 58 = 7.76, p = 0.007). Given the extreme values for both response types in the Levene's test, the mixed ANOVA was decided not to be performed. The reason for this was that the extremely low levels of p values in the Levene's test indicate a clear-cut violation of the homogeneity of variances assumption, and increase the likelihood of Type I error rate, leading to false positives. Hence, a significant effect could likely be resulting from a false positive, and mislead the overall results. A non-parametric test was performed instead, which showed a significant difference between the confidence levels of R and K judgments ($\chi 2(1) = 35.1$, $p < 0.001$; see Figure 3.2)

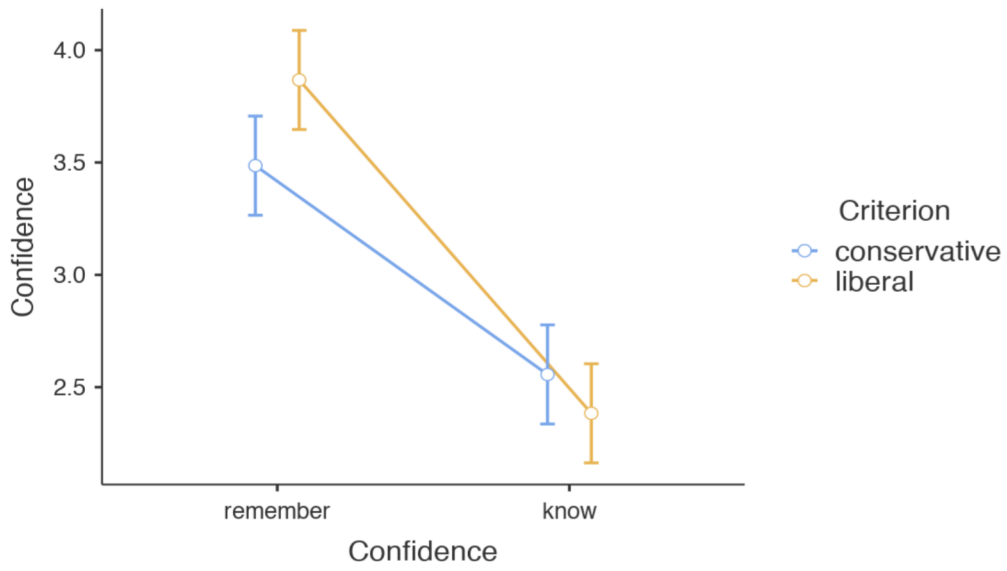

Figure 3.2 Confidence levels of R and K responses, grouped by conservative and liberal response criteria.

3.2.3.4. d' and C Values for R and K Responses

A repeated measures ANOVA was performed on the sensitivity (*d'*) scores of R and K responses. The sensitivity scores for the R and K responses were calculated based on the HR and FAR calculations of the two response types, as explained above. The sphericity and homogeneity assumptions were partially met (Levene's F (1, 58) = 0.794, p = 0.377 in *d'* for R; F (1, 58) = 5.005, p = 0.029 in d' for K). The mixed ANOVA was still performed to examine the differing values of *d'* for r and K responses as a function of criterion.

The results revealed a significant difference between the *d'* scores (F (1, 35) = 327.85, MSE = 70.621, $p < 0.001$, $\eta 2p$ = 0.850; see Figure 2.5) where R responses (M = 2.223, SE = 0.087, *95% CI* = [2.050, 2.396]) were made with greater sensitivity than K responses (M = 0.688, SE = 0.087, *95% CI* = [0.515, 0.861]). Moreover, there was a significant effect of criterion (F (1, 58) = 33.9, MSE = 23.508, $p < 0.001$, $\eta 2p$ = 0.369), and a criterion by sensitivity interaction (F (1, 58) = 7.35, MSE = 1.582, $p < 0.01$, $\eta 2p$ = 0.112). The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that sensitivity for R responses were greater than the sensitivity for K responses in both conservative (t (58) = 10.71, MD = 1.305, SE = 0.122, $p < 0.001$) and liberal conditions (t (58) = 14.97, MD = 1.765, SE = 0.118, $p < 0.001$; see Table 3.2).

Table 3.2

d' values for R and K Responses in Experiment 2

				95% Confidence Interval	
dPrime	Criterion	Mean	SE	Lower	Upper
remember	conservative	2.551	0.124	2.305	2.796
	liberal	1.895	0.123	1.651	2.138
know	conservative	1.245	0.124	1.000	1.491
	liberal	0.130	0.123	−0.114	0.374

As for the criterion values of the R and K responses, HR and FARs for the respective response types were utilized for the calculation of the *C* scores, similar to the calculation of *d'*. The results of the repeated measures ANOVA revealed a nonsignificant difference between the *C* values of R and K responses (F (1, 58) = 0.34, MSE = 0.066, p = 0.562, $\eta 2p$ = 0.006). The interaction between *C* values of R and K responses and criterion was also nonsignificant (F (1, 58) = 0.001, MSE = 0.0002, p = 0.972, $\eta 2p$ = 0.000). However, the effect of criterion on the corresponding C values was significant (F (1, 58) = 27.6, MSE = 6.048, p < 0.001, $\eta 2p$ = 0.322; see Table 3.3), mimicking the results of the between-subjects test.

Table 3.3

C values for R and K Responses in Experiment 2

				95% Confidence Interval	
Criterion	C	Mean	SE	Lower	Upper
conservative	remember	0.898	0.0836	0.732	1.064
	know	0.848	0.0836	0.683	1.014
liberal	remember	0.446	0.0823	0.283	0.609
	know	0.402	0.0823	0.239	0.565

3.2.3.5. Hits and False Alarms for K Responses

In order to compare the mean hit and false alarm values of the K responses as a function of criterion, a mixed ANOVA was performed on the corresponding number of hits and false alarms for the K responses. The K hits and K false alarms were set as the repeated measures factor, and the criterion was set as the between subject factor. The sphericity assumption was met due to the repeated measure having only two levels (hits and false alarms). Levene's test of homogeneity of variances was met for the K false alarms (F (1, 58) = 3.05, p = 0.086), but not for K hits (F (1, 58) = 9.20, p = 0.004). However, considering the small sample size of the data set as well as the data being raw, the mixed ANOVA was still performed given the robustness of the test.

According to the results of the ANOVA, there was a significant difference between K hits and K false alarms (F (1, 58) = 74.4, MSE = 18840.2, $p < 0.001$, $\eta 2p$ = 0.562). Moreover, there was a significant interaction between K responses of hits and false alarms, and criterion (F (1, 58) = 34.1, MSE = 8464.2, $p < 0.001$, $\eta 2p$ = 0.370). There was also a significant effect of criterion on the K hits and false alarms (F (1, 58) =

39.3, *MSE* = 18288.2, $p < 0.001$, $\eta 2p = 0.404$) The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that there were more hits than false alarms in the liberal condition (t (58) = 10.404, *MD* = 49.51, *SE* = 4.88, $p < 0.01$). However, the difference between hits and false alarms was not significant in the conservative condition (t (58) = 1.935, *MD* = 8.00, *SE* = 4.14, $p = 0.348$). The overall K-hits were more than the overall K false alarms (t (58) = 8.62, *MD* = 24.8, *SE* = 2.88, $p < 0.01$). Moreover, there were more hits than false alarms in the liberal condition (t (58) = 6.27, *MD* = 24.7, *SE* = 3.94, $p <$ 0.01; see Table 3.4 and Figure 3.3).

Table 3.4

Number of Hits and False Alarms for K Responses in Experiment 2

				95% Confidence Interval	
Criterion	KNOW	Mean	SE	Lower	Upper
conservative	Hit	29.0	3.47	22.1	35.9
	False Alarm	21.0	3.47	14.1	27.9
liberal	Hit	70.5	3.43	63.7	77.3
	False Alarm	28.9	3.43	22.1	35.7

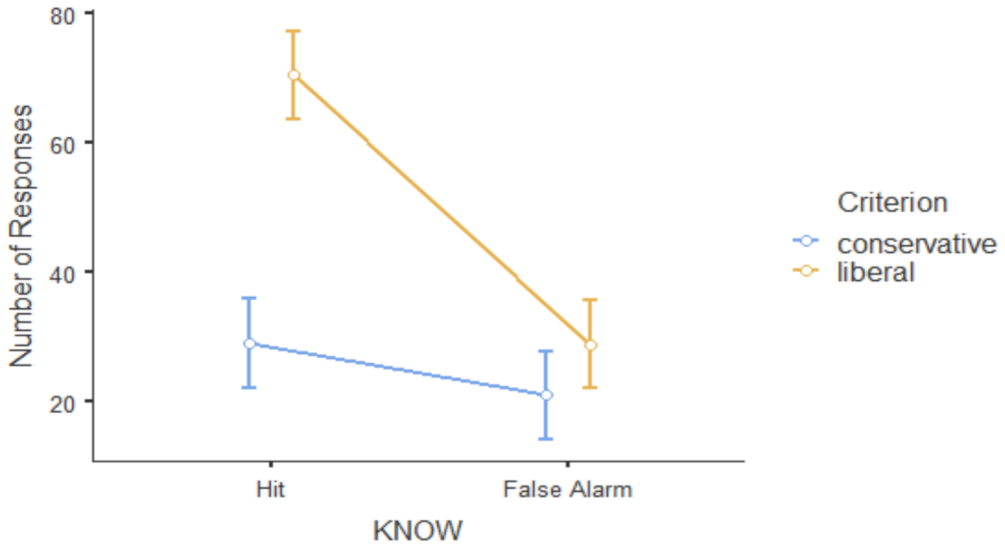

Figure 3.3 Hits and false alarms for K responses in Experiment 2.

3.2.3.6. Hits and False Alarms for R Responses

A mixed ANOVA was performed on the number of hits and false alarms for the R responses. The R hits and R false alarms were set as the repeated measures factor, and the criterion was set as the between subject factor. The sphericity assumption was met due to the repeated measure having only two levels (hits and false alarms). Levene's test of homogeneity of variances was met for the R false alarms (F (1, 58) = 0.470, p = 0.496), but not for R hits (F (1, 58) = 4.358, p = 0.041). However, considering the small sample size of the data set as well as the data being raw, the mixed ANOVA was still performed given the robustness of the test.

Table 3.5

Number of Hits and False Alarms for R Responses in Experiment 2

				95% Confidence Interval	
Criterion	REMEMBER	Mean	SE	Lower	Upper
conservative	Hit	45.12	3.92	37.345	52.9
	False Alarm	7.60	3.92	−0.172	15.4
liberal	Hit	107.53	3.86	99.889	115.2
	False Alarm	8.89	3.86	1.244	16.5

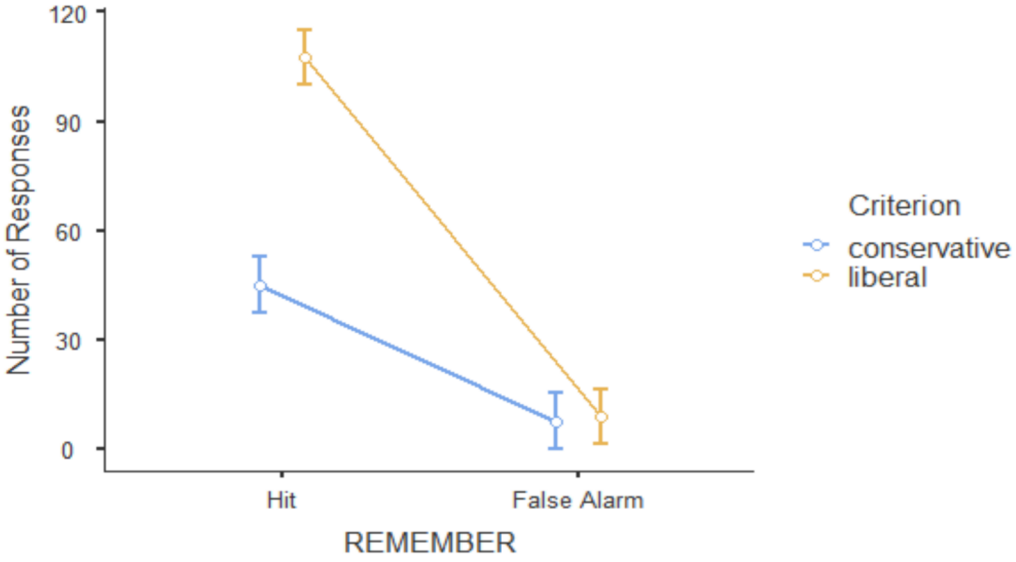

Figure 3.4 Hits and false alarms for R responses in Experiment 2.

According to the results of the ANOVA, there was a significant difference between R hits and R false alarms (F (1, 58) = 302.6, MSE = 1138897, $p < 0.001$, $\eta 2p = 0.839$). Moreover, there was a significant interaction between R responses of hits and false alarms, and criterion (F (1, 58) = 61.0, MSE = 27994, $p < 0.001$, $\eta 2p = 0.513$). There was also a significant effect of criterion on the R hits and false alarms (F (1, 58) = 67.7, MSE = 30399, $p < 0.001$, $\eta 2p = 0.539$) The post-hoc tests were carried out using Bonferroni adjustment for p values. The results of the post-hoc comparisons revealed that there were more hits than false alarms in both liberal (t (58) = 18.128, MD = 98.65, SE = 5.44, $p < 0.01$) and conservative conditions (t (58) = 16.669, MD = 37.52, SE = 5.63, $p < 0.01$). The overall R-hits were more than the overall R false alarms (t (58) = 17.4, MD = 68.1, SE = 3.91, $p < 0.01$). Finally, the difference between hit and false alarms was greater in the liberal condition than it was in the conservative condition (t (58) = 8.23, MD = 31.8, SE = 3.87, $p < 0.01$).

3.2.3.7. Hit Rates and False Alarm Rates for K Responses

In order to compare the mean HRs and FARa of the K responses as a function of criterion, a mixed ANOVA was performed on the corresponding HRs and FARs for the K responses. The “know” HRs and FARs were set as the repeated measures factor, and the criterion was set as the between subject factor. The sphericity assumption was met due to the repeated measure having only two levels (hit rates and false alarm rates). However, Levene's test of homogeneity of variances was not met either for the HRs (F (1, 58) = 18.5, $p < 0.001$), or the FARs (F (1, 58) = 26.1, $p < 0.001$). Hence, the mixed ANOVA was not performed due to the violation of the homogeneity assumption. Instead, for comparison purposes, a non-parametric repeated measures ANOVA was performed on the HRs and FARs for K responses. This way, the results could be compared with the results from the number of hits and false alarms. There was a statistically significant difference in the HRs and FARs, $\chi 2(1) = 34.3$, $p < 0.001$ (Table 3.6).

Table 3.6

HRs and FARs for K Responses in Experiment 2

Criterion	Response Rate	Mean	SE	95% Confidence Interval	
				Lower	Upper
conservative	HR	0.4421	0.0269	0.3886	0.496
	FAR	0.0946	0.0269	0.0411	0.148
liberal	HR	0.3771	0.0267	0.3240	0.430
	FAR	0.3321	0.0267	0.2790	0.385

3.2.3.8. Hit Rates and False Alarm Rates for R Responses

A mixed ANOVA was performed on the corresponding HRs and FARs for the R responses. The "remember" HRs and FARs were set as the repeated measures factor, and the criterion was set as the between subject factor. The sphericity assumption was met due to the repeated measure having only two levels (hit rates and false alarm rates). However, Levene's test of homogeneity of variances was not met either for the HRs (F (1, 58) = 9.33, p = 0.003), or the FARs (F (1, 58) = 26.05, p < 0.001).

Hence, the mixed ANOVA was not performed due to the violation of the homogeneity assumption. Instead, a non-parametric repeated measures ANOVA was performed on the HRs and FARs for K responses. There was a statistically significant difference in the HRs and FARs, $\chi2(1) = 60.0$, $p < 0.001$ (Table 3.7).

Table 3.7

HRs and FARs for R Responses in Experiment 2

				95% Confidence Interval	
Criterion	Response Rate	Mean	SE	Lower	Upper
conservative	HR	0.6310	0.0256	0.5802	0.6818
	FAR	0.0342	0.0256	−0.0166	0.0850
liberal	HR	0.6837	0.0252	0.6337	0.7336
	FAR	0.1322	0.0252	0.0822	0.1821

3.3. Discussion

Experiment 2 revealed results that are similar to the results of Experiment 1 with a few differences. First, the results of Experiment 2 revealed that the sensitivity between the conservative and liberal conditions was not the same. The responses in the conservative condition pointed to a greater sensitivity as compared to the liberal condition. The criterion value *C*, however, was greater in the conservative condition, indicating that the intended criterion shift occurred between the conditions. Similar to the differing sensitivity levels of the conservative and liberal conditions, the sensitivity associated with R responses was greater than the sensitivity of K responses. This implies a parallel between the low and high thresholds of criteria, and familiarity and recollection response thresholds. On a similar note, the *C* value associated with R responses was higher than the *C* value associated with K responses, drawing a similar parallel. In addition, R responses were made with greater confidence than K responses. The second difference between the first and the second experiments was that the number of R and K responses did not differ from one another as they did in Experiment 1. Although the number R and K responses did not differ significantly, there was a greater difference between them in the liberal condition, shaped by a larger number of R responses than K responses. The proportion of responses were not significantly different under

different criteria either, further suggesting that R and K response tendencies may not be affected by response bias.

For the K responses, the number of hits was greater than the number of false alarms in the liberal condition. However, the difference was not significant in the conservative condition. As for the R responses, there was a greater number of hits in both conditions, the difference being greater in the liberal condition. The repetitive finding that the number of hits is higher than the number of false alarms suggests the possibility that the experiment was too easy. When we look at *d'* as a performance measure, we can see that there are especially high *d*'s in R answers. Regarding the hit and false alarm rates, the data set did not form a normal distribution with equal variances. This may be due to the transformation of raw data from frequencies to proportions of responses. As discussed in the previous chapter, the unequal distribution of targets and foils tend to inflate or deflate the respective hit and false alarm rates, resulting in a larger difference between hit and false alarm rates. Hence, the results are interpreted with a grain of salt until better comparisons can be made.

CHAPTER 4

GENERAL DISCUSSION

In the current study, the decision criterion was attempted to be shifted through base rate manipulations based on the assumptions of a signal detection model. In accordance with this purpose, the corresponding familiarity and recollection judgments were examined in relation to the differing criteria. The confidence ratings of the respective responses were also taken, and analyzed in the experiments. The results suggested that familiarity and recollection judgments differ from one another in terms of lower and higher decision criteria. The results of the experiments are discussed in relation to similar studies in the literature, limitations of the current study are considered, and potential future directions are mentioned.

4.1. Discussion

The results indicated a direct effect of the criterion change on *C* scores, and a partial effect on sensitivity. The effect of criterion on *C* values suggests that participants moved their response bias in the intended direction when the criterion was liberal or conservative. In both experiments, participants in the conservative condition had higher *C* values that were above a neutral value of 0, whereas participants in the liberal condition had lower *C* values below 0. This means that participants who were in the conservative condition required more information to have made an "old" judgment whereas participants in the liberal condition required less information to make an "old" judgment. As for sensitivity scores, the two experiments yielded different results. It

was hypothesized that *d'* would not differ among different criteria because sensitivity is assumed to be independent of criterion (Stanislaw & Todorov, 1999). Although this assumption was met in Experiment 1 with *d'* values that are not significantly different under different criteria, a higher *d'* was observed in the conservative condition in Experiment 2. The situations in which *d'* may vary across conditions has been discussed in the literature in terms of various conditions. For instance, metacognitive monitoring and performance feedback has been suggested to improve performance during a recognition test (Selmeczy & Dobbins, 2013). In this context, how well participants incorporate internal evidence, such as the trial-to-trial gradations of internal memory evidence (i.e. metacognitive monitoring ability) during decision making can be a determining factor in participants' performance. In Experiment 2, *d'* was greater in the conservative condition than it was in the liberal condition, and it was greater for recollection judgments than it was for familiarity judgments. These results can be interpreted in terms of a higher engagement of metacognitive monitoring associated with a higher criterion, which results in less false alarms. This can be further deducted from the low frequency of false alarms in Experiment 2. The fact that 12 participants have 0, and 7 participants have only 1 false alarm for their R responses should be taken into consideration in terms of the difference between hit and false alarms, and an indicator of high performance. There is also evidence suggesting that there is a gradual decrease in memory performance as a function of test block when there is a criterion manipulation (Pala, 2019). Accordingly, the criterion shift gets more prominent with the increasing number of trials, which can be observed from the increasing differences in C values between consequent blocks. This suggests that participants require a certain number of trials and/or blocks before they can set their criterion in the desired direction.

In terms of the number and the proportion of R and K responses, the results indicated a partial effect of base rate and criterion on the frequency R and K responses. This effect was partial in the sense that it was only observed in Experiment 1 when a base rate manipulation was applied to the test list. In experiment 1, more R responses were

given in the liberal, and more K responses were given in the conservative conditions. Consistent with this finding, R responses were made with higher confidence than K responses. When put together, these two findings paint a picture of gradual confidence steps on a memory strength continuum. When the criterion is liberal, less evidence is required to make an R judgment, and since a lower threshold meets these needs, more R judgments are made with confidence. However, when the criterion is conservative, there needs to be more evidence in order for R judgments to be made, which results in less R responses than K responses. The same line of results were not found in Experiment 2, in which the number of R and K responses did not differ from one another under different criteria. The proportion of R and K responses did not differ in either of the experiments. Moreover, although the frequency of familiarity and recollection judgments did not differ from one another in Experiment 2, the difference in confidence levels remained. R responses were made with more confidence than K responses. This finding is consistent with other findings in the literature, which support a single-process model of familiarity and recollection based on a signal detection model (Dunn, 2004; Stretch & Wixted, 1998; Wixted, 2007; Wixted & Stretch, 2004).

In terms of the hits and false alarms, both the frequency and the proportion of these response groups were calculated and analyzed in the experiments. However, the calculation of hit and false alarm rates for the R and K judgments in the corresponding criterion conditions revealed some drawbacks, which complicated the interpretation of the result. The standard calculation of HRs and FARs is the number of hits or false alarms, divided by the number of targets or lures, respectively. However, two issues occur when calculating HRs and FARs for only the "old" judgments that are under different base rates. First, the fact that there are no misses or correct rejections associated with R and K responses means that their hit and false alarm rates will be smaller than expected because they represent a small sample of responses from a larger pool (i.e. They represent a sub-group of "old" responses among all targets or all lures. For instance, there may never be an R or K hit rate equal to 1 because the total of the R or K responses are always less than the total number of targets). This is because not

every item gets a chance of being judged on the basis of familiarity and recollection, only old items do. As a result of this, the denominator gets the usual number of total targets or lures but the numerator is always taken from a smaller sample of possible items. The fact that these rates are smaller than expected, had they been calculated based on the total of the "old" responses, may not cause concern as the same deflation will apply to both R and K responses. However, a notable concern ensues when these rates are calculated based on different criterion conditions. Thus, the second issue that occurs is that the HRs and FARs for responses under conservative and liberal conditions are biased in the following way: Conservative HRs are inflated because the denominator (i.e. number of targets) is smaller due to a 30% target rate, whereas the conservative FARs are deflated because the denominator (i.e. number of lures) is larger due to a 70% lure rate.

The opposite is true for the liberal condition, in which HRs are deflated and FARs are inflated. This trend is most clearly seen in the boosted HRs of both K and R responses in Experiment 1. For the K responses, the graph completely changes the direction, and for the R responses it gets steeper with the inflation. However, even when the frequency of hits and false alarms are taken as the basis, the results indicate that there are more hits than false alarms for both response types when the condition is liberal. This is contradicting with the predictions of Donaldson's (1996) two-threshold model, in which, the FAR is expected to be greater than the HR for familiarity judgments when the response bias is liberal. The contradicting results may be associated with the overall performance in the task characterized by a high number of hits and a low number of false alarms. In other words, task difficulty might have interfered with the hits and false alarms in a way that prevents any sound interpretation of the data.

4.2. Limitations

In this thesis, familiarity and recollection was examined under two different criteria in a standard recognition memory task. The obtained data were analyzed based on

criterion, base rate and response type variables. The division of data in a mixed design requires dividing the data set into different groups, and given the small sample size of the current study, the groups had either 15 or 30 participants each. Hence, the small sample size increased the effect of outliers on the overall group means, and resulted in non-normal data distributions in many instances, which might have affected the current findings. Moreover, a between-subject design was preferred for the experimental design which might have created the following issues. First, the between-subject design suggested that each participant acted as their own control in terms of criterion setting, and that there was no neutral criterion with which the liberal or conservative criteria could have been compared. Following a block design and setting the criterion gradually with increasing or decreasing base rates could have provided a more controlled criterion manipulation.

One of the main findings that contradicted the existing literature was the higher number of hit rates as compared to false alarms, especially for K responses in the liberal condition. According to the SDT model of familiarity and recollection judgments, a liberal criterion indicates a threshold that is closer to the noise distribution as compared to a conservative one. Accordingly, a consequence of the liberal criterion on the two-threshold model is that since both K and R thresholds are set more to the left of the distribution, the area between them consists of more false alarms than hits. The current findings may suggest that the criterion manipulation was not strong enough to create the predicted outcome, which is difficult to tell as specific criterion values for R and K responses are inflated due to non-normal HRs and FARs. However, given this possibility, a lower/higher proportion of targets can be used for the conservative (e.g. 20%) and liberal (e.g. 80%) conditions to make a more pronounced criterion shift. On a similar note, the number of trials in each block can be increased in order to increase the difficulty of the task, and have more false alarms and less hits.

4.3. Future Research

The future directions related to the current study are discussed in terms of methodological and model-based implications. First, as mentioned above, a within-subjects design can be applied with more blocks that create a gradual shift into the desired criterion in future studies examining the effect of criterion on familiarity and recollection judgments. In addition to the within-subjects design, different methods of criterion manipulations can be applied alone, or in combination, in order to obtain a stronger criterion shift. Among these methods, base rates of test lists, pay-offs, memory strength of items, response feedback, and prior information can be included (Curran, DeBuse,& Leynes, 2007; Han & Dobbins, 2008, 2009; Van Zandt, 2000). The behavioral studies examining familiarity and recollection in relation to response criteria are relatively rare when compared to the studies using physiological methods such as EEG or fMRI (Azimian-Faridani, & Wilding, 2006; Brezis, Bronfman, Yovel & Goshen-Gottstein, 2017; Smith, Wixted & Squire, 2011; Woroch & Gonsalves, 2010). This suggests that although there is extensive research and theory available in this area, there is little methodological consensus on the procedural nature of the behavioral studies. This situation highlights the importance of behavioral and methodological studies in this field. Future studies may delve more into the use of the R-K procedure, as well as the measurement and calculation of the related variables that are examined along with these judgments.

Second, future research may involve modelling studies to examine which recognition memory model best explains *remember-know* data. The implication of these results with respect to other models of recognition memory can be discussed in terms of different models of recognition memory such as the source of activation confusion (SAC) model, which supports a dual-process explanation of familiarity and recollection, and the sum-difference theory of remembering and knowing (STREAK; Rotello, Macmillan & Reeder, 2004) and the univariate signal detection (Wixted & Stretch, 2004) models, which support a single-process explanation (see Starns &

Ratcliff, 2008 for a comparison of the two models). Among these models, the SAC model predicts the proportion of R and K responses that that would be produced under different conditions in a recognition memory task. These prediction values are based on the activation values of the memory traces within the model. A single-process model, STREAK incorporates SDT model parameters such as *d'*, and *C* to explain the characteristics of the memory evidence. The univariate signal detection model is based on the assumptions of the standard SDT model, and assumes that familiarity and recollection judgments reflect levels of confidence. In this context, the results obtained in the current study support a single-process model of familiarity and recollection based on the fact that R and K judgments represent high and low confidence responses, as well as high and low thresholds. However, the lack of a significant effect on the basis of HRs and FARs in line with the assumptions of the single-process models makes it insufficient to explain familiarity and recollection processes with single-process models. Hence, the current data do not fully support either of the models based on these findings. Nevertheless, given the methodological directions mentioned above, future studies can be conducted with a particular aim to collect data suitable for model testing to be able to vouch for a specific model. As a result, the obtained data are informative in the sense that they provide a methodological blueprint to be followed in a similar study to be carried out for the purpose of model fitting in the future. It will especially be useful in studies that will examine criterion manipulation and R-K answers, considering the large effect size of the current study's criterion indicator *C* under different criteria.

4.4. Conclusion

The findings of the current study was in a different direction than the expected results based on the existing literature in some parts, yet revealed important results in terms of familiarity and recollection processes which broaden our knowledge on recognition memory. It was found that familiarity and recollection judgments not only reflect low and high confidence responses, they also reflect low and high

thresholds on the memory strength continuum. This supports the two-threshold SDT model of recognition memory according to which, “remember” and “know” responses are not independent from each other. In terms of the hits and false alarm rates associated with these responses, better designs and measurements may be required in future studies to better examine how the two processes would behave under different response biases. These results provide evidence that support the single-process model of recognition memory, and further suggest that familiarity and recollection can be expressed in terms of liberal and conservative decision processes. The finding of the contradicting results may help us develop more reliable methods in measuring these judgments under criterion shifts, and develop better models of recognition memory.

REFERENCES

Azimian-Faridani, N., & Wilding, E. L. (2006). The influence of criterion shifts on electrophysiological correlates of recognition memory. *Journal of Cognitive Neuroscience, 18*(7), 1075-1086.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74(2)*, 81-99.

Brezis, N., Bronfman, Z. Z., Yovel, G., & Goshen-Gottstein, Y. (2017). The electrophysiological signature of remember–know is confounded with memory strength and cannot be interpreted as evidence for dual-process theory of recognition. *Journal of cognitive neuroscience, 29*(2), 322-336.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42*(8), 1088-1106.

Curran, T., DeBuse, C., & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of experimental psychology: Learning, Memory, and Cognition, 33*(1), 2.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*(4), 523-533.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological review, 111*(2), 524.

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, 94*(11), 5973-5978.

Ebbinghaus, H. (1913). Memory (ha ruger & ce bussenius, trans.). *New York: Teachers College.(Original work published 1885)*, *39*.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new'effects: memory strength and decisional factor (s). *Neuropsychologia, 40*(13), 2288-2304.

Gardiner, J. M., Gregg, V. H., & Karayianni, I. (2006). Recognition memory and awareness: Occurrence of perceptual effects in remembering or in knowing depends on conscious resources at encoding, but not at retrieval. *Memory & cognition, 34*(2), 227-239.

Gardiner, J. M., & Richardson-Klavehn, R. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). Oxford, England: Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: Evidence for adaptive criterion learning. *Memory & Cognition, 36*(4), 703-715. doi:10.3758/MC.36.4.703

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review, 16*(3), 469-474. doi:10.3758/PBR.16.3.469

Hayes, B. K., Dunn, J. C., Joubert, A., & Taylor, R. (2017). Comparing single-and dual-process models of memory development. *Developmental Science, 20*(6), e12469.

Herrmann, Douglas J. and Roger Chaffin, 1988, *Memory in Historical Perspective: The Literature Before Ebbinghaus*, New York: Springer.

Hirshman, E. (1995). Decision processes in recognition memory: criterion shifts and the list strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 302.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science, 9*(1), 61-65.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306.

James, W. (1890). *Principles of psychology*. New York: Holt Knott, L. M., & Dewhurst, S. A. (2007). Divided attention at retrieval disrupts knowing but not remembering. *Memory, 15*(6), 664-674.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207* (4427), 203-205.

Koop, G. J., Criss, A. H., & Malmberg, K. J. (2015). The role of mnemonic processes in pure-target and pure-foil recognition memory. *Psychonomic Bulletin & Review, 22*(2), 509-516.

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic bulletin & review*, *3*(2), 164-170.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah. NJ, US: Lawrence Erlbaum Associates Publishers.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review, 87*(3), 252.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14*(5), 858-865.

Milner, B. (1962). Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérales. In *Physiologie de l'hippocampe* (pp. 257–272). Paris: Centre National de la Recherche Scientifique. English translation: B. Milner & S. Glickman (Eds.). Princeton: Van Nostrand, 1965 (pp. 97–111).

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: l4 year follow-up study of H.M. *Neuropsychologia, 6*, 215–234.

Murdock, B. B., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology, 94*(3), 284-290.

Murre, J. M., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PloS one*, *10*(7), e0120644.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review, 110* (4), 611-46.

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General, 139*(2), 341–364. https://doi.org/10.1037/a0018926

Opitz, B. (2010). Context-dependent repetition effects on recognition memory. *Brain and cognition, 73* (2), 110-118.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience, 9*, 1465–1472.

Pala, C. D. (2019). *The Roles of Feedback and Prior Information on the Effectiveness of the Base Rate Manipulation for Response Bias* (thesis).

Paller, K., Kutas, M., & Mayes, A. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67* (4), 360-371.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Science, 6* (2), 93-102.

R Core Team (2018). R: A Language and environment for statistical computing. [Computer software]. Retrieved from https://cran.r-project.org/.Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Science, 11* (6), 251-7.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychological Review, 111*(3), 588.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392* (6676), 595-598.

Ryle, G. (1949). *The concept of mind.* San Francisco, CA: Hutchinson.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology, 17*(6), 568-576.

Schacter, D., Chiu, C., & Ochsner, K. (1993). Implicit memory: A selective review. *Annual review of neuroscience, 16* (1), 159-182.

Schmaltz, L. W., & Theios, J. J. (1972). Acquisition and extinction of a classically conditioned response in hippocampectomized rabbits (Oryctologaus cuniculus). *Journal of Comparative and Physiological Psychology, 79,* 328–333.

Shimamura, A. P., & Squire, L. R. (1987). A neuropsychological study of fact memory and source amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 464-473.

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1499.

Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memories are strong. *Journal of Neuroscience, 31*(44), 15693-15702.

Spanton, R. W., & Berry, C. J. (2020). The unequal variance signal-detection model of recognition memory: Investigating the encoding variability hypothesis. *Quarterly Journal of Experimental Psychology, 73*(8), 1242-1260.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory, 82*(3), 171-177.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers, 31*(1), 137-149.

Starns, J. J., & Ratcliff, R. (2008). Two dimensions are not better than one: STREAK and the univariate signal detection model of remember/know performance. *Journal of memory and language, 59*(2), 169-182.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based andfrequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379.

Tekcan, A. I., & Göz, I. (2005). Türkçe kelime normları (Turkish word norms). Istanbul: Bogazici Universitesi Yayinevi.

The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software]. Retrieved from https://www.jamovi.org.

Tulving, E., Schacter, D., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent on recognition memory. *Learning, Memory and Cognition, 8*, 336–341.

Tulving, E. (1984). Precis of elements of episodic. *Behav Brain Sci*, *7*(2), 223-38.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.

Tulving, E. (1993). What is episodic memory?. *Current directions in psychological science*, *2*(3), 67-70.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 301–306.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 582-600. doi:10.1037/0278-7393.26.3.582

Voss, J., & Paller, K. (2009). Remembering and knowing: Electrophysiological distinctions at encoding but not retrieval. *Neuroimage, 46* (1), 280-289.

Warren, L. R. (1980). Evoked potential correlates of recognition memory. *Biological Psychology, 11*(1), 21-35.

Warrington, E. K., & Weiskrantz, L. (1970). The amnesic syndrome: Consolidation or retrieval?. *Nature, 228*, 628–630.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616-641.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review, 114*(1), 152.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological review, 117*(4), 1025.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, 11(4), 616-641.

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research, 1100*, 125–135.

Woroch, B., & Gonsalves, B. (2010). Event-related potential correlates of item and source memory strength. *Brain Research, 1317* , 180-191.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of experimental psychology: Learning, memory, and cognition, 20*(6), 1341.

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*(3), 323.

Yonelinas, A. P. (1999). Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams (1999).

Yu, S. Y., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research, 1320*, 74–84.

APPENDICES

A. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816/ 01 ARALIK 2021

Konu : Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi : İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Aslı Kılıç ÖZHAN

Danışmanlığını yürüttüğünüz Zeynep Erbaş'ın "Bellekte Karar Verme" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **494-ODTU-2021** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

Prof.Dr. Mine MISIRLISOY
İAEK Başkanı

B. THE INFORMED CONSENT FORM

ARAŞTIRMAYA GÖNÜLLÜ KATILIM FORMU

Bu çalışma ODTÜ Psikoloji Bölümü öğretim üyelerinden Dr. Aslı Kılıç Özhan tarafından yürütülmektedir. Bu form sizi araştırma koşulları hakkında bilgilendirmek için hazırlanmıştır.

Çalışmanın Amacı Nedir?

Bu çalışma yeni bilgiler öğrenirken belleğimize bu bilgileri nasıl kaydettiğimizi ve daha sonra nasıl hatırladığımızı araştırmaktadır.

Bize Nasıl Yardımcı Olmanızı İsteyeceğiz?

Araştırma Psikoloji Bölümü Bilişsel Psikoloji Laboratuvarı'nda yapılacaktır. Üniversite öğrencileri katılımcı olarak davet edilecek, katılmak isteyenler yaklaşık 1 saatlik bir laboratuvar seansına katılacaklardır. Çalışmada size kelimeler gösterilecektir. Daha sonra bu kelimeleri hatırlayıp hatırlamadığınız sorulacaktır.

Katılımınızla ilgili bilmeniz gerekenler:

Bu çalışmaya katılmak tamamen gönüllülük esasına dayalıdır. Herhangi bir yaptırıma veya cezaya maruz kalmadan çalışmaya katılmayı reddedebilir veya çalışmayı bırakabilirsiniz. Araştırma esnasında cevap vermek istemediğiniz sorular olursa boş bırakabilirsiniz.

Araştırmaya katılanlardan toplanan veriler tamamen gizli tutulacak, veriler ve kimlik bilgileri herhangi bir şekilde eşleştirilmeyecektir. Katılımcıların isimleri bağımsız bir listede toplanacaktır. Ayrıca toplanan verilere sadece araştırmacılar ulaşabilecektir. Bu araştırmanın sonuçları bilimsel ve profesyonel yayınlarda veya eğitim amaçlı kullanılabilir, fakat katılımcıların kimliği gizli tutulacaktır.

Çalışmaya katılanlar bu duyurunun yapıldığı ders için puan alacaklardır. Alınacak puan dersin öğretim üyesi tarafından belirlenecektir.

Riskler:

Çalışma ile ilgili bilinen bir risk yoktur.

Araştırmayla ilgili daha fazla bilgi almak isterseniz:

Çalışmayla ilgili soru ve yorumlarınızı araştırmacıya odtu.bellek@gmail.com adresinden iletebilirsiniz.

Yukarıdaki bilgileri okudum ve bu çalışmaya tamamen gönüllü olarak katılıyorum.
(Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyad Tarih İmza

---/----/-----

C. THE INSTRUCTIONS OF THE EXPERIMENT

Çalışmamıza katıldığınız için teşekkür ederiz. Bu çalışma, hafızamızdaki bilgilere nasıl eriştiğimizi araştırmaktadır. Bu deney 4 bloktan oluşmakta ve her blok çalışma, aritmetik ve test listelerini içermektedir. Dilerseniz bloklar arasında kısa aralar verebilirsiniz.

Çalışma bölümünde bilgisayar ekranının ortasında arka arkaya kelimeler göreceksiniz. Sizden, bu kelimeleri çalışıp daha sonra hatırlamanız beklenmektedir.

Çalışma listesinin ardından ekranda sırayla rakamlar göreceksiniz. Bu bölümde göreviniz, çıkan rakamları birbirleri üzerine ekleyerek toplamak. Mesela ilk rakam gösterildiğinde cevap olarak gösterilen rakamı yazacaksınız (ör, 5). Ardından gelen ikinci rakamı (ör, 2) bir önceki cevabınıza ekleyeceksiniz. Bu durumda cevabınız ilk iki rakamın toplamı olacak (ör, toplam: 7). Cevabınızı, klavyenin üst kısmındaki rakamları kullanarak yalnızca cevabınız olan sayıyı yazıp ENTER tuşuna basarak verebilirsiniz. Bu şekilde bu bölüm bitene kadar devam edeceksiniz.

Toplama işlemini bitirdiğinizde test bölümü başlayacak. Test bölümünde bilgisayar ekranının ortasında arka arkaya kelimeler sunulacak. Test boyunca göreceğiniz kelimelerin %30'u daha önce çalıştığınız kelimelerden oluşacak. Geriye kalan kelimeler ise daha önce görmediğiniz, yeni kelimelerden oluşacak. Sizden yapmanız istenen, her bir kelimenin daha önceki çalışma listesinde olup olmadığına karar vermek. Eğer kelimenin daha önceki çalışma listesinde olduğunu düşünüyorsanız "Eski", yeni bir kelime olduğunu düşünüyorsanız "Yeni" cevabını klavyedeki "z" ve "m" tuşlarını kullanarak verebilirsiniz. Lütfen cevabınızı vermeden önce test kelimesinin ekrandan gitmesini bekleyin. Cevabınızı, ekranda "Eski" ve "Yeni" kelimeleri belirdiği zaman verebilirsiniz.

Daha sonra sizden verdiğiniz cevaptan ne kadar emin olduğunuzu 1 ve 4 arasında bir sayıyla belirtmeniz istenmektedir. Cevabınızı klavyenin üst kısmındaki rakamları kullanarak yazıp ENTER tuşuna basarak verebilirsiniz. Eğer test kelimesinin daha önceki çalışma listesinde olduğuna karar verdiyseniz, sizden bu test kelimesine dair ikinci bir karar vermeniz istenmektedir. Kelimenin daha önceki çalışma listesinde olduğunu biliyor, ama kelimeyi çalıştığınız anı hatırlamıyorsanız "Biliyorum"; kelimenin daha önceki çalışma listesinde olduğunu biliyor ve kelimeyi çalıştığınız anı hatırlıyorsanız "Hatırlıyorum" cevabını, klavyedeki "s" ve "j" tuşlarını kullanarak verebilirsiniz.

Yolda yürürken karşıdan gelen birini gördünüz ve bu kişinin tanıdığınız biri olduğuna karar verdiniz. Bu kişiyi tanıdığınızdan ve yüzünü bildiğinizden emin olsanız da bir türlü bu kişinin adını, kim olduğunu, bu kişiyle ne zaman veya nerede tanıştığınızı hatırlamıyorsunuz. Bu durumda bu kişiyi biliyor ama hatırlamıyorsunuz. Daha sonra, bu kişiyle geçen sene birlikte İngilizce dersi aldığınız aklınıza geliyor, adını ve kim olduğunu fark ediyorsunuz. Ne zaman ve nerede tanıştığınız aklınıza geliyor. Bu durumda bu kişiyi hatırlıyorsunuz.

Şimdi yapacağımız deneyde de, benzer bir şekilde, sizden test sırasında eski olduğunu düşündüğünüz kelimeler için o kelimeyi biliyor musunuz, yoksa hatırlıyor musunuz karar vermenizi istiyorum. Eğer eski olduğuna karar verdiğiniz bir kelimeyi çalıştığınız ana geri dönebiliyorsanız “HATIRLIYORUM’, eski olduğuna karar verdiğiniz bir kelimeyi çalıştığınız ana geri dönemiyorsanız ‘BİLİYORUM” cevabını verebilirsiniz.

Her test kelimesinden önce sol işaret parmağınızı "z" tuşunun üzerinde, sağ işaret parmağınızı ise "m" tuşunun üzerinde tutmaya özen gösterin. Deneye başlamadan önce pratik yapmanız için kısa bir alıştırma bloğu olacak. Daha sonra 4 bloktan oluşan kısım başlayacak.

Hazırsanız alıştırmaya başlamak için BOŞLUK tuşuna basabilirsiniz.

D. TURKISH SUMMARY / TÜRKÇE ÖZET

BÖLÜM 1

GİRİŞ

1.1. Tanıma Belleği

İnsan belleği bildirimsel ve bildirimsel olmayan bellek olarak ikiye ayırır; bunlardan birincisi ile hatırlama yoluyla bilinçli olarak geri getirilebilen olayların temsilleri, ikincisi ile ise önceden öğrenilmiş hafıza izlerinin aktivasyonu yoluyla geri çağrılan, performansla ilgili anılar kastedilir. Bildirimsel bellek ayrıca episodik ve semantik bellek olarak ikiye ayrılır (Tulving, 1984, 1993). Episodik bellek, olaylarla ilgili anıları oluştururken semantik bellek, bağlamsal detaylar olmaksızın bilgilerle ilgili anıları oluşturur. Episodik ve semantik bellek, belirli bellek sistemlerine girdiği varsayılan bellek görevleriyle incelenebilir. Özellikle episodik bellek, tanıma belleği görevlerinin uygulandığı laboratuvar ortamlarında yaygın olarak araştırılmaktadır.

Standart bir tanıma belleği görevinde madde tanıma testi, katılımcıların bir madde listesi (kelimeler, resimler vb.) üzerinde çalışması ve daha sonra bu maddeler üzerinden test edilmesiyle uygulanır. Çalışma listesi, katılımcılardan çalışmaları istenen bir dizi maddeden oluşur. Herhangi bir tekrar veya bir yenilik etkisini (yani, en son çalışılan maddelerin daha sonra geri çağrılma şansının daha yüksek olduğu bir seri konum etkisi) önlemek için genellikle çalışma ve test aşamaları arasında kısa bir çeldirici görev vardır (Ebbinghaus, 1913/1885; Murre & Dros, 2015). Test listesi, hedeflerden (önceden çalışılan maddeler) ve folyolardan (yeni maddeler) oluşur. Katılımcıların testteki görevi, hedefleri folyolardan ayırmaktır. Hafıza performansı,

hedef öğelere verilen eski veya evet yanıtlarının oranı (isabet oranı) ve yanlışlara verilen eski veya evet yanıtlarının oranı (yanlış alarm oranı) ile ölçülür.

1.2. Sinyal Tespit Teorisi

Sinyal tespit teorisi (SDT; Green & Swets, 1966; Banks, 1970) modellerine göre tanıma belleği görevi sırasında test maddelerine dair verilen eski ve yeni kararları, eski ve yeni öğeler için bellek gücü dağılımlarıyla eşleştirilebilen bir güven spektrumuna dayanmaktadır. Bu bağlamda eski ve yeni kelimeler iki farklı dağılımla temsil edildiği ve bu dağılımların normal olarak dağıldığı varsayılır. Bununla birlikte eski madde dağılımı için olan ortalama bellek gücü, yeni madde dağılımı için ortalama bellek gücünden daha büyüktür. Karar verme sırasında katılımcılar bu dağılımlar üzerinde bir yerde bir yanıt ölçütü belirler ve ilgili bellek izinin gücünün yanıt ölçütünü aşıp aşmadığına göre bir tanıma kararı verilir. Bir maddenin bellek gücü yanıt ölçütünü aşarsa "eski" yanıt verilir, maddenin bellek gücü yanıt ölçütünün altındaysa "yeni" yanıt verilir.

Kriter, aşinalık düzeyini ölçütü aşan maddelerin eski, aşinalık düzeyi ölçütün altında kalan maddelerin yeni olarak tanımlandığı ve katılımcılar tarafından belirlenen bir değerdir. Dolayısıyla ölçüt değerinin yerleşimi bellek yargılarını etkileyen faktörlerden biridir. Kriter çizgisinin tarafsız, yani eski ve yeni kararlarının eşit şanslarla verildiği, konumun soluna çizilen daha liberal bir kriter, daha fazla isabet ve daha fazla yanlış alarm oranı ile ilişkilendirilir. Buna karşılık, kriter çizgisinin tarafsız konumun sağına çizilen daha muhafazakar bir kriter, daha az isabet ve yanlış alarm oranı ile ilişkilendirilir.

1.3. Tanıma Belleğinde tek ve İkili Süreç Modelleri

Tanıma belleği deneylerinden elde edilen bulguların açıklanmasında ağırlıklı olarak iki model öne sürülmüştür: Tek süreç ve ve ikili süreç modelleri. Bu modeller arasında tek süreç modelleri, madde tanıma kararlarının, bellek gücü ve karar faktörlerinden

etkilenen tek bir sürece dayandığını öne sürmektedir (Donaldson, 1996; Wixted & Stretch, 2004; Wixted, 2007). Bu anlamda tek süreç modelleri, eski ve yeni ayrımının ötesindeki herhangi bir ikili veya çoklu bellek yargısının, bellek gücü sürekliliği üzerindeki farklı eşikler temelinde açıklanabileceğini ve tanıma belleği kararlarının sinyal tespit kuramı bağlamında ele alınabileceğini önermektedir.

Tek süreç modellerinin aksine ikili süreç modelleri, madde tanıma kararlarının niteliksel olarak farklı iki bellek sürecine dayandığını öne sürer (Mandler, 1980). Bu modellere göre aşinalık ve hatırlama süreçleri, bir öğenin tanınıp tanınmadığına ilişkin karara birlikte katkıda bulunur. Özellikle aşinalığın bir maddenin bellek gücüne karşılık gelen sürekli bir süreç olduğu, hatırlamanın ise ya hep ya hiç şeklinde gerçekleşen bir eşik süreci olduğu ileri sürülmektedir (Onyper, Zhang ve Howard, 2010; Yonelinas, 1994). Bu anlamda ikili süreç modelleri, aşinalık ve hatırlama yargılarının sürekli süreçler olan aşinalık ve hatırlamaya ortak bir paydada dayandığını savunan tek süreç modellerine aksi yöndedir

1.4. Bilme ve Hatirlama

Tek süreç ve ikili süreç modellerini karşılaştırmak adına çeşitli davranışsal çalışmalar yapılmıştır. İkili süreç modellerini destekleyen çalışmalarda, literatürde bilme/hatirlama paradigması olarak bilinen bir prosedür yaygın olarak uygulanmaktadır (Tulving, 1984, 1993). Katılımcıların yanıtlarını varsayılan çerçeveye göre yönlendirmek için prosedür, katılımcıların test maddesinin episodik bellekten alınıp alınmadığına dair bir hatırlama kararı vermelerini ve test maddesinin semantik bellekten alınıp alınmadığına dair bir bilme kararı vermelerini gerektiriyordu. Bu doğrultuda katılımcılardan test maddesi ve bu maddeye dair bağlamsal ayrıntıları hatırlıyorlarsa “hatırlıyorum” yanıtını vermeleri, test maddesinin daha önce sunulduğunu bildikleri halde bağlamsal detayları hatırlamıyorlarsa “biliyorum” yanıtını vermeleri istenmiştir. Daha yakın zamanlarda bu yöntem, episodik bellekte aşinalık ve hatırlamanın bir ölçüsü olarak kullanılmıştır. Bilme/hatırlama paradigması ile elde edilen birçok bulgu, bilme ve hatırlama

cevaplarının tekabül ettiği aşinalık ve hatırlama süreçleri arasında bir ayrışmanın olduğunu düşündürmektedir (Gardiner, Gregg ve Karayianni, 2006; Gardiner, Ramponi ve Richardson-Klavehn, 1998; Gardiner ve Richardson-Klavehn, 2000; Knott & Dewhurst , 2007; Yönelinas ve diğerleri, 1998). Bununla birlikte bu bulgular, aşinalık ve hatırlama yargılarının bir bellek gücü sürekliliği üzerindeki kademeli güven seviyelerini yansıttığını öne süren tek süreç modelleri tarafından aksi yönde tartışmalarla eleştirilmiştir (Wixted, 2007). Bu nedenle, ikili süreç modellerinin savunduğu ve fonksiyonel bir ayrışmaya işaret eden herhangi bir kanıt dizisinin, önerilen manipülasyonlar doğrultusunda modelin ilgili parametrelerinin değiştirilmesiyle sinyal tespit teorisi modeli ile açıklanabileceği tartışılmaktadır.

1.5. Donaldson Modeli

Donaldson (1996) tarafından önerilen modelde bilme ve hatırlama cevaplarının sinyal tespit kuramı bağlamında ifade edilen eski ve yeni madde dağılımları üzerine yerleştirilen iki kriter eşiği ile ifade edilebileceği öne savunulmuştur . Bu bağlamda birinci eşiğin bir maddenin eski veya yeni olduğunu belirleyen sınır, ikinci eşiğin ise maddenin bilinip bilinmediğini veya hatırlanıp hatırlanmadığını belirleyen sınır olduğu varsayılan bir iki eşikli sinyal tespit modeli öne sürülmüştür. Dolayısıyla, bu modele göre üç olası yanıt vardır. İlk olarak bir maddenin bellek gücü ilk eşiğin altına düşerse yeni olarak tanımlanır. İkinci olarak maddenin bellek gücü eğer iki eşik arasına düşerse eski olarak tanımlanır ve "bilme" cevabıyla ilişkilendirilir. Son olarak maddenin bellek gücü ikinci kriterin üzerinde ise madde eski olarak tanımlanır ve bir hatırlama cevabıyla ilişkilendirilir.

Bu kriterlerin eski ve yeni dağılımları üzerinde yerleştirilebileceği yerler karar yanlılığına göre değişir. Örneğin, tbir maddenin eski olduğu kararının %50 şans oranından daha yüksek bir oranla verildiği liberal bir kriter kabul edilebilir. Alternatif olarak, yeni yanıtının verme olasılığının daha yüksek olduğu ve eski yanıtının daha nadir verildiği muhafazakar bir kriter benimsenebilir. Üç olası kriter yerleşiminin değerlendirilmesi, bir tanıma belleği görevi sırasında eski/yeni ve bilme/hatırlama

yanıtlarının oranları hakkındaki performans tahminlerini şu şekilde ortaya çıkarır: İlk olarak, (1) duyarlılık parametresi (yani d') bellek performansının ölçütsüz bir tahminini sağlar; bu nedenle, "hatırlama" yanıtları için d', genel tanıma performansı için olan d'ye eşit olmalıdır. İkincisi, (2) üç olası kriter koşulu (yani liberal, tarafsız ve muhafazakar kriterler) dikkate alındığında, "bilme" yanıtları farklı kriterler altında eski/yeni yanıt eşiğinden bağımsız değildir. Buna göre "bilme" yanıtları incelendiğinde, muhafazakar bir kriter altında "bilme" kararına ait isabet oranı yanlış alarm oranından daha yüksek, tarafsız kriterler altında eşit ve liberal kriterler altında daha düşük olmalıdır. Ayrıca, d' muhafazakar kriterler altında en büyük değerini alır, tarafsız kriterlerde 0'a eşittir ve liberal kriterlerde 0'dan küçüktür. Bu tahminlerin arkasındaki mantık, ilgili üç koşul için, dağılımlar üzerindeki eşikler yer değiştirdiğinde eski ve yeni dağılımlarının bu eşikler arasında kalan alanlarına bakarak anlaşılabilir.

1.6. Tezin Amacı

Mevcut çalışmada, aşinalık ve hatırlamanın sinyal tespit teorisine dayanan tek süreç modeli, kriter değişiminin "bilme" ve "hatırlama" yargıları üzerindeki etkilerini inceleyen iki deneyde araştırılmaktadır. Bu doğrultuda, "bilme" ve "hatırlama" yargılarının yüksek ve düşük eşik yanıtlarının yanı sıra yüksek ve düşük güven yanıtlarını yansıtması beklenmektedir. Ayrıca Donaldson'ın modelinin (1996) öngörülerine uygun olarak, eski/yeni ve hatırlama eşikleri arasındaki alanın “bilme” yanıtlarını temsil ettiği göz önüne alındığında, isabet ve yanlış alarm oranları açısından belirli sonuçlar beklenmektedir. İlk olarak, kriter muhafazakar olduğunda isabet oranı yanlış alarm oranından daha yüksek olmalı, kriter liberal olduğunda ise yanlış alarm oranından daha düşük olmalıdır. İkincisi d', hafızanın kriterden bağımsız bir tahminini sağlamalıdır, yani her kriter koşulunda da eşit olmalıdır. Ancak Donaldson, liberal koşulun dağılımında işbet oranının yanlış alarm oranından daha az olması nedeniyle bellek performansının liberal kriter altında şansın altında olabileceğini savunuyor. Dolayısıyla 0'dan küçük bir d' değeri, modele göre hala beklenen değerler içindedir. Bu tahminlerle uyumlu bir sonuç modeli, Donaldson'ın modeli açısından aşinalık ve

hatırlama yargılarının tek süreçli bir model olduğu yorumunu destekleyecektir. Ancak, tek süreç modeline alternatif ikili süreç yorumlarının, özellikle Donaldson'ın isabet ve yanlış alarm oranlarına dair katı tahminleri dikkate alındığında, bu modelden farklı olacaği da göz önünde bulundurulmalıdır. Tek süreç modellerinin aksine, aşinalık ve hatırlama yargıları arasında kriter veya güven değerlendirmelerine dayanan bir ilişkinin olmaması, iki sürecin birbirinden bağımsız olduğunu öne sürecek ve bu da ikili süreç modellerini destekleyecektir.

BÖLÜM 2

DENEY

2.1. Katılımcılar

Birinci deneyde araştırmaya Orta Doğu Teknik Üniversitesi'nden (ODTÜ) 40 lisans öğrencisi katılmıştır. Yapılan deneyler ODTÜ İnsan Denekleri Etik Kurulu tarafından onaylanmıştır. Öğrenciler, katılımları karşılığında kısmı ders kredisi almıştır. Tüm katılımcılar, deney başlamadan önce yazılı onay vermiştir.
Kötü performans ve teknik aksaklık sebebiyle verinin kullanılamaması sonucu 36 katılımcıdan toplanan (O yaş = 22.4, SD yaş = 3.75) veriler kullanılmıştır. Katılımcı havuzu 7 erkek ve 29 kadın, 3 solak ve 33 sağ elini kullanan kişiden oluşmuştur. Tüm katılımcılar, normal renk görüşüne ve normal veya normale göre düzeltilmiş görme keskinliğine sahip ve anadili Türkçe olan kişilerdir.

İkinci deneyde ise çalışmaya Orta Doğu Teknik Üniversitesi'nden (ODTÜ) 62 lisans öğrencisi katılmıştır. Öğrenciler, katılımları karşılığında kısmi ders kredisi almış ve deney başlamadan önce yazılı onay vermiştir. İki katılımcıdan alınan veriler kötü performans sebebiyle son analizden çıkarılmıştır. Bu nedenle, iki katılımcının verisi

çıkarıldıktan sonra, analizde 60 katılımcının (O yaş = 21.8, SD yaş = 1.83) verileri kullanılmıştır. Katılımcıların 10'u erkek ve 52'si kadındır. 7 katılımcı solak ve 55 katılımcı ise baskın olarak sağ elini kullanmaktadır. Tüm katılımcılar, normal renk görüşüne ve normal veya normale göre düzeltilmiş görme keskinliğine sahip ve anadili Türkçe olan kişilerdir.

2.2. Uyaranlar

Sözcükler, tanıma belleği deneyinde kullanılmak üzere Türkçe Sözcük Normlarından (Tekcan ve Göz, 2005) rastgele seçilmiştir. Deneyde kullanılan kelimeler dört ila dokuz harf arasında değişmektedir. 1232 kelimenin tam listesi, daha sonra deneyin MATLAB betiğinde kullanılmak üzere bir metin belgesine yazılmış ve deney MATLAB'da Psychtoolbox uzantısı kullanılarak kodlanmıştır.

2.3. Desen ve Prosedür

Birinci deneyde kişiler arası iki faktör yer almaktadır (temel oran: sabit veya manipüle edilmiş ve kriter: liberal veya muhafazakar). Katılımcılar her koşula rastgele atanmıştır. Deney süresince, 10 çalışma denemesi ve 10 test denemesinden oluşan bir pratik bloğu ve her biri 80 çalışma denemesi ve 80 test denemesinden oluşan dört deney bloğu vardı. Kelime listesi her oturumdan önce karıştırıldı ve her katılımcıya rastgele sunuldu.

Temel oran manipülasyonu için, katılımcıların yarısına test maddelerinin %30'unun çalışılan maddelerden oluşacağı bir test listesi, diğer yarısına ise test maddelerinin %70'inin çalışılan maddelerden oluşacağı bir test listesinin sunulacağı söylendi. Aslında her gruptaki katılımcıların yalnızca yarısında bir temel oran manipülasyonu yapıldı. Diğer yarısı, eşit sayıda hedef ve folyo içeren bir liste ile test edildi.

%30 temel oran koşulu muhafazakar koşulu, %70 temel oran koşulu ise liberal koşulu oluşturmuştur. Buna göre, %30 temel oran koşulundaki katılımcıların daha katı bir

kriter belirlemeleri ve test maddelerine "yeni" yanıt verme olasılıklarının daha yüksek olması bekleniyordu. Öte yandan, %70 temel oran koşulundaki katılımcıların daha geniş bir ölçüt belirlemeleri ve test maddelerine "eski" yanıt verme olasılıklarının daha yüksek olması bekleniyordu.

İkinci deneyde birinci deneydekini prosedürün aynısı uygulanmış, yalnızca temel oran koşulu çıkarılarak her katılımcıya temel oranı değiştirilmiş bir test listesi sunulmuştur. Aynı zamanda katılımcılar verdikleri her yanıttan sonra doğru veya yanlış şeklinde bir geri bildirim almış, böylelikle performansın yükselmesi amaçlanmıştır.

2.4. Sonuçlar

2.4.1. Veri Transformasyonu

Duyarlılık (d') değerleri, kriter (C) değerleri, isabet oranı ve yanlış alarm oranı, aşağıdaki istatistiksel analizlerden önce her koşul için hesaplanmıştır. İsabet oranı (İO), toplam isabet sayısının toplam hedef sayısına bölünmesiyle hesaplanmıştır. Yanlış alarm oranı (YAO) ise toplam yanlış alarm sayısının toplam folyo sayısına bölünmesiyle hesaplanmıştır. Yanıt yanlılığı ölçüsü C ve duyarlılık ölçüsü d', her koşul için isabet ve yanlış alarm oranları kullanılarak hesaplanmıştır.

$$d' = z(HR) - z(FAR)$$

$$C = -0.5[z(HR) - z(FAR)]$$

İO (H) = "Hatırlama" sayısı / ("Hatırlama" sayısı + Toplam kaçırma sayısı)

İO (B) = "Bilme" sayısı / ("Bilme" sayısı + Toplam kaçırma sayısı)

YAO (H) = "Hatırlama" yanlış oran sayısı / ("Hatırlama" yanlış oran sayısı + Toplam doğru ret sayısı)

YAO (B) = "Bilme" yanlış oran sayısı / ("Bilme" yanlış oran sayısı + Toplam doğru ret sayısı)

d' (H) = z (İO [h]) - z (YAO [h])

d' (B) = z (İO [b]) - z (YAO [b])

*C (H) = -0.5 * [z (İO [h]) + z (YAO [h])]*

*C (B) = -0.5 * [z (İO [b]) + z (YAO [b])]*

2.4.2. Deney 1

Temel oran ve kriterin d' değerleri üzerindeki etkisini araştırmak için iki yönlü bir ANOVA gerçekleştirilmiştir. Temel oranın duyarlılık üzerindeki etkisi ($F (1, 32) = 0.030$, $MSE = 0.010$, $p = 0.864$, $\eta 2p = 0.001$)veya kriterin duyarlılık üzerindeki ana etkisi ($F (1, 32) = 1.726$, $MSE = 0.588$, $p = 0.198$, $\eta 2p = 0.051$) anlamlı bulunmamıştır. Temel oran ve kriter arasında da benzer bir şekilde bir etkileşim etkisi bulunmamıştır ($F (1, 32) = 0.020$, $MSE = 0.007$, $p = 0.889$, $\eta 2p = 0.001$).

Temel oran ve kriterin C değerleri üzerindeki etkisini araştırmak için iki yönlü bir ANOVA gerçekleştirilmiştir. Kriterin yanıt yanlılığı üzerindeki etkisi anlamlı bulunmuştur ($F (1, 32) = 6.839$, $MSE = 0.477$, $p < 0.05$, $\eta 2p = 0.176$). Muhafazakar kriter koşulu, beklendiği gibi liberal kriter koşulundan ($M = -0.143$, $SE = 0.062$) daha büyük bir C değeri ($M = 0.087$, $SE = 0.062$) ile ilişkilendirilmiştir (Tablo 2.3). Temel oranın yanıt yanlılığına etkisi anlamsız bulunmuştur ($F (1, 32) = 0.174$, $MSE = 0.012$, $p = 0.680$, $\eta 2p = 0.005$) ve temel oran ve kriter arasındaki etkileşim de anlamlı çıkmamıştır ($F (1, 32) = 0,260$, $MSE = 0,018$, $p = 0,613$, $\eta 2p = 0,008$).

Kriterin ($F (1, 32) = 0.890$, $MSE = 0.010$, $p = 0.352$, $\eta 2p = 0.027$) ve temel oranın ($F (1, 32) = 0.055$, $MSE = 0.001$, $p = 0.815$, $\eta 2p = 0.002$) isabet ve yanlış alarm oranları

üzerindeki etkisi anlamsız bulunmuştur. Aynı şekilde etkileşim etkisi de anlamsızdır ($F(1, 32) = 0.258$, $MSE = 0.003$, $p = 0.615$, $\eta 2p = 0.008$). Dolayısıyla, ortaya çıkan isabet ve yanlış alarm oranları üzerinde ne kriterin ne de temel oran koşullarının bir etkisi olmuştur.

Sonuçlar, kriterin "bilme" ve "hatırlama" yanıtları üzerinde anlamlı bir etkisi ($F(1, 32) = 34.667$, $MSE = 14706.13$, $p < 0.001$, $\eta 2p = 0.520$) ve temel oran ve kriter etkileşimi ($F(1, 32) = 6.719$, $MSE = 2850.13$, $p < 0.05$, $\eta 2p = 0.174$) olduğunu göstermiştir. Post-hoc testler, p değerleri için Bonferroni ayarlaması kullanılarak gerçekleştirilmiştir. Post-hoc karşılaştırmaların sonuçları, yalnızca temel oran manipülasyonu olduğunda muhafazakar ve liberal koşullar arasında "bilme" ve "hatırlama" yanıtları arasındaki ortalama farkın anlamlı olduğunu ortaya koymuştur ($t(32) = 6.00$, $MD = 41.2$, $SE = 6.87$, $p < 0{,}001$).

Buna ek olarak "bilme" ve "hatırlama" yanıtlarının ($F(1, 31) = 54.860$, $MSE = 20.477$, $p < 0.001$, $\eta 2p = 0.639$) güven seviyeleri arasında önemli bir değişiklik olduğunu görülmüştür. Buna göre "hatırlama" yanıtları "bilme" yanıtlarından daha yüksek güvenle verilmiştir. Ancak kriter etkileşimine göre güven ($F(1, 31) = 0.046$, $MSE = 0.017$, $p = 0.832$, $\eta 2p = 0.001$), temel oran etkileşimine göre güven ($F(1, 31) = 0.214$, $MSE = 0.080$, $p = 0.647$, $\eta 2p = 0.007$), veya kriter ve temel oran etkileşiminin anlamlı olmadığı ($F(1, 31) = 2.758$, $MSE = 1.029$, $p = 0.107$, $\eta 2p = 0.082$) bulunmuştur. Buna göre güven puanları, yanıt yanlılığından ve temel oran manipülasyonlarından bağımsız olarak birbirlerinden ayrışmaktadır.

ANOVA'nın sonuçları, "bilme" cevaplarına ait isabet ve yanlış alarm oranları arasında önemli bir fark göstermiştir ($F(1, 32) = 156.876$, $MSE = 2.249$, $p < 0.001$, $\eta 2p = 0.831$). Ayrıca, isabet oranı ve yanlış alarm oranlarının "bilme" yanıtları ile kriter ($F(1, 32) = 5.205$, $MSE = 0.075$, $p < 0.05$, $\eta 2p = 0.140$) arasında anlamlı bir etkileşim bulunmuştur. "Bilme" yanıtları ve temel oranı arasındaki etkileşim anlamlı bulunmamıştır ($F(1, 32) = 0.460$, $MSE = 0.007$, $p = 0.503$ $\eta 2p = 0.014$), "bilme" yanıtları, kriter ve temel oranı arasındaki etkileşim de bilakis anlamlı değildir (F (1,

32) , 32) = 1.285, MSE = 0.018, p = 0.265, η2p = 0.039). Fakat isabet ve yanlış alarm oranları üzerinde de kriterin anlamlı bir etkisi vardır (F (1, 32) = 6.666, MSE = 0.126, p < 0.05, η2p = 0.172). Post-hoc karşılaştırmaların sonuçları, temel oran koşulundan bağımsız olarak (t (32) = 2.58, MD = 0.08, SE = 0.032, p < 0.05) kriter liberal olduğunda isabet oranının yanlış alarm oranından daha büyük olduğunu ortaya koymuştur.

"Hatırlama" cevapları açısından isabet ve yanlış alarm oranları arasında anlamlı bir fark bulunmuştur (F (1, 32) = 291.016, MSE = 6.346, p < 0.001, η2p = 0.901).Kriterin "hatırlama" cevaplarına ait isabet oranı ve yanlış alarm oranı üzerindeki etkisi isabet oranlarının daha yüksek olduğu bir grafik çizmiştir (F (1, 32) = 4.635, MSE = 0.105, p < 0.05, η2p = 0.127). Post-hoc karşılaştırmaların sonuçları, temel orandan bağımsız olarak (t (32) = 17.1, MD = 0.59, SE = 0.03, p < 0.001) her iki kriter koşulunda da isabet oranlarının yanlış alarm oranlarından daha büyük olduğunu ortaya koymuştur.

2.4.3. Deney 2

Kriterin duyarlılık üzerindeki etkisi anlamlı bulunmuştur (F (1, 58) = 31.8, MSE = 10.342, p < 0.001, η2p = 0.354). d' ile ölçülen duyarlılık, muhafazakar koşulda (M = 1.99, SE = 0.106, %95 CI = [1.778, 2.20]) liberal durumda olduğundan (M = 1.16, SE = 0.102, %95 CI = [0.954, 1.36]) daha yüksektir. Bu, amaçlanan kriter manipülasyonunun, sinyal tespit teorisinde varsayıldığı gibi, bir kriter kayması yoluyla tanıma testi sırasında katılımcıların tepki yanlılığında bir kayma sağladığını göstermiştir. Buna göre C, muhafazakar koşulda daha yüksektir (M = 0.224, SE = 0.043, %95 CI = [0.138, 0.311]) ve liberal koşula kıyasla daha katı bir kriteri gösterir (M = -0.249, SE = 0.042, %95 GA = [-0.333, -0.166]).

İsabet oranı açısından, kriter etkisi önemsiz bulunmuştur (F (1, 58) = 3.13, MSE = 0.05, p = 0.082, η2p = 0.050). Ancak yanlış alarm oranları üzerinde kriterin önemli bir etkisi olduğu görülmüştür (F (1, 58) = 51.0, MSE = 0.93, p < 0.001, η2p = 0.460). Post-hoc karşılaştırmalar yanlış alarm oranlarının liberal durumda muhafazakar

durumda olduğundan daha yüksek olduğunu ortaya koymuştur (t (58) = 7.14, MD = 0.245, SE = 0.03, $p < 0.001$). Bunun haricinde iki yönlü ANOVA'nın sonuçları, kriterin (F (1, 58) = 0.090, MSE = 0.022, p = 0.765, $\eta 2p$ = 0.002) karşılık gelen güven seviyeleri üzerinde önemli bir etkisi olmadığını göstermiştir.

Sonuçlar, kriterin "bilme" ve "hatırlama" yanıtları üzerinde önemli bir etkisi olduğunu göstermiştir (F (1, 58) = 545, MSE = 95844, p < 0,001, $\eta 2p$ = 0.904). Post-hoc karşılaştırmanın sonucu, liberal koşulda "bilme" ve "hatırlama" yanıtları arasındaki ortalama farkın muhafazakar durumda olduğundan daha büyük olduğunu ortaya koymuştur (t (58) = 23.3, MD = 56.6, SE = 2.42, $p < 0.001$). "Bilme" ve "hatırlama" yanıtlarına ait güven seviyeleri arasında yapılan parametrik olmayan test sonucunda anlamı bir etki bulunmuştur ($\chi 2(1) = 35.1$, $p < 0.001$).

Duyarlılık (d') değerleri arasında da (F (1, 35) = 327.85, MSE = 70.621, $p < 0.001$, $\eta 2p$ = 0.850) arasında anlamlı bir fark bulunmuştur. Buna göre "hatırlama" yanıtları (M = 2.223, SE = 0.087, %95 GA = [2.050, 2.396]), "bilme" yanıtlarından daha yüksek hassasiyetle verilmiştir (M = 0.688, SE = 0.087, %95 GA = [0.515, 0.861]). Ayrıca, kriterin anlamlı etkisi (F (1, 58) = 33.9, MSE = 23.508, $p < 0.001$, $\eta 2p = 0.369$). Aynı zamanda "bilme" ve "hatırlama" yanıtlarının C değerleri (F (1, 58) = 0.34, MSE = 0.066, $p = 0.562$, $\eta 2p = 0.006$) arasında anlamlı olmayan bir fark ortaya çıkmıştır.

Son olarak "bilme" isabetleri ile "bilme" yanlış alarmları arasında anlamlı bir fark bulunmuştur (F (1, 58) = 74.4, MSE = 18840.2, $p < 0.001$, $\eta 2p$ = 0.562). Ayrıca, "bilme" isabetleri ile yanlış alarmlar ve kriter (F (1, 58) = 34.1, MSE = 8464.2, $p < 0.001$, $\eta 2p = 0.370$) arasında anlamlı bir etkileşim bulunmuştur. Bu doğrultuda yanlış alarmlardan daha fazla isabet vardır (t (58) = 10.404, MD = 49.51, SE = 4.88, $p < 0.01$).

"Bilme" isabetleri ile "hatırlama" yanlış alarmları arasında anlamlı bir fark vardır (F (1, 58) = 302.6, MSE = 1138897, $p < 0.001$, $\eta 2p$ = 0.839). Ayrıca, "hatırlama" isabetleri ile yanlış alarmları ve kriter (F (1, 58) = 61.0, MSE = 27994, $p < 0.001$, $\eta 2p$

= 0.513) arasında anlamlı bir etkileşim vardır. Hem liberal (t (58) = 18.128, MD = 98.65, SE = 5.44, $p < 0.01$) hem de muhafazakar koşullarda (t (58) = 16.669, MD = 37.52, SE = 5.63, $p < 0.01$) yanlış alarmlardan daha fazla isabet vardır.

"Bilme" cevaplarına ait isabet oranı ve yanlış alarm oranı açısından yapılan parametrik olmayan testler sonucunda istatistiksel olarak anlamlı bir fark bulunmuştur, $\chi2(1) = 34.3$, $p < 0.001$. Benzer bir şekilde "hatırlama" cevaplarına ait isabet ve yanlış alarm oranlarında da yapılan parametrik olmayan testte anlamlı bir etki görülmüştür $\chi2(1) = 60.0$, $p < 0.001$.

BÖLÜM 3

GENEL TARTIŞMA

3.1.Genel Tartışma

Sonuçlar, kriter değişikliğinin C puanları üzerindeki doğrudan etkisine ve duyarlılık üzerinde kısmi bir etkiye işaret etmiştir. Kriterin C değerleri üzerindeki etkisi, kriter liberal veya muhafazakar olduğunda, katılımcıların yanıt yanlılıklarını amaçlanan yönde hareket ettirdiğini göstermektedir. Her iki deneyde de, muhafazakar koşuldaki katılımcılar nötr bir değer olan 0'ın üzerinde daha yüksek C değerlerine sahipken, liberal koşuldaki katılımcılar 0'ın altında daha düşük C değerlerine sahipti. Bu, muhafazakar durumdaki katılımcıların “eski” bir yargıda bulunmak için daha fazla bilgiye ihtiyaç duyduğu, liberal durumdaki katılımcıların ise “eski” bir yargıda bulunmak için daha az bilgiye ihtiyaç duyduğu anlamına gelir. Duyarlılık puanlarına gelince, iki deney farklı sonuçlar vermiştir. Duyarlılığın kriterden bağımsız olduğu varsayıldığından, d'nin farklı kriterler arasında farklılık göstermeyeceği varsayılmıştır (Stanislaw ve Todorov, 1999). Bu varsayım, Deney 1'de farklı kriterler altında önemli

ölçüde farklı olmayan d' değerleri ile karşılansa da, Deney 2'de konservatif koşulda daha yüksek bir d' gözlemlenmiştir.

Sonuçlar, “bilme” ve “hatırlama” yanıtlarının frekansı üzerinde temel oran ve kriterin kısmi bir etkisi olduğunu gösterdi. Bu etki, yalnızca Deney 1'de test listesine bir temel oran manipülasyonu uygulandığında gözlemlenmesi anlamında kısmiydi. Deney 1'de, liberal koşulda daha fazla “hatırlama” yanıtı ve muhafazakar koşulda daha fazla “bilme” yanıtı verildi. Bu bulguyla tutarlı olarak, “hatırlama” yanıtları, “bilme” yanıtlarından daha yüksek bir güvenle yapılmıştır. Bir araya getirildiğinde, bu iki bulgu, bir hafıza gücü sürekliliği üzerinde kademeli güven adımlarının resmini çiziyor. Kriter liberal olduğunda, bir “hatırlama” yargısı yapmak için daha az kanıt gerekir ve daha düşük bir eşik bu ihtiyaçları karşıladığından, daha fazla “hatırlama” yargısı güvenle yapılır. Bununla birlikte, kriter muhafazakar olduğunda, “hatırlama” yargılarının yapılabilmesi için daha fazla kanıt olması gerekir, bu da “bilme” yanıtlarından daha az “hatırlama” yanıtıyla sonuçlanır. Aynı sonuç, “bilme” ve “hatırlama” yanıtlarının sayısının farklı kriterler altında birbirinden farklı olmadığı Deney 2'de bulunamamıştır. Fakat aşinalık ve hatırlama yargılarının sıklığı birbirinden farklı olmasa da güven düzeylerindeki fark devam etmiştir. “Hatırlama” yanıtları, “bilme” yanıtlarından daha güvenle yapılmıştır. Bu bulgu, sinyal tespit teorisine dayalı tek süreçli aşinalık ve hatırlama modelini destekleyen literatürdeki diğer bulgularla tutarlıdır (Dunn, 2004; Stretch & Wixted, 1998; Wixted, 2007; Wixted & Stretch, 2004).

3.2.Sınırlamalar

Bu tezde küçük örneklem boyutu, aykırı değerlerin genel grup ortalamaları üzerindeki etkisini arttırmış ve normal olmayan veri dağılımlarına neden olmuştur. Ayrıca, mevcut literatürle çelişen bazı bulgular mevcuttur. Bunlardan biri, liberal koşuldaki "bilme" cevapları için yanlış alarmlara kıyasla daha yüksek isabet oranlarıydı. Sinyal tespit teorisi modeline göre liberal koşulda hem "bilme" hem de "hatırlama" eşikleri

dağılımın daha soluna ayarlandığından, aralarındaki alan isabetlerden daha fazla yanlış alarmdan oluşur.

Mevcut bulgular, kriter manipülasyonunun öngörülen sonucu yaratmak için yeterince güçlü olmadığını göstermektedir. Bu olasılık göz önüne alındığında, daha belirgin bir kriter değişikliği yapmak için muhafazakar (ör. %20) ve liberal (ör. %80) koşullar için daha düşük/yüksek bir hedef oranı kullanılabilir. Benzer bir şekilde görevin zorluğunu artırmak ve daha fazla yanlış alarm ve daha az isabet olması için her bloktaki deneme sayısı artırılabilir.

3.3.Gelecek Çalışmalar

Kriterin aşinalık ve hatırlama yargıları üzerindeki etkisini inceleyen gelecekteki çalışmalarda, istenen kritere kademeli bir geçiş yaratan ve daha fazla blok kullanılan bir denek içi desen uygulanabilir. Denek içi desene ek olarak daha güçlü bir kriter değişimi elde etmek için farklı kriter manipülasyon yöntemleri tek başına veya kombinasyon halinde uygulanabilir. Buna ek olarak gelecekteki araştırmalar, hangi tanıma belleği modelinin "hatırlama" ve "bilme" verilerini daha iyi açıkladığını incelemek için modelleme çalışmalarını içerebilir. Bu sonuçların diğer tanıma belleği modellerine göre anlamı, farklı tanıma belleği modelleri açısından tartışılabilir.

3.4.Sonuç

Mevcut çalışmanın bulguları aşinalık ve hatırlama yargılarının yalnızca düşük ve yüksek güvenli cevapları yansıtmadığını, aynı zamanda bellek gücü sürekliliği üzerindeki düşük ve yüksek eşikleri de yansıttıklarını bulmuştur. Bu, “hatırlama” ve “bilme” yanıtlarının birbirinden bağımsız olmadığına göre iki eşikli ve sinyal tespit teorisi odaklı tanıma belleği modelini destekler. Bu sonuçlar, tek süreçli tanıma belleği modelini destekleyen kanıtlar sağlar ve ayrıca aşinalık ve hatırlamanın liberal ve muhafazakar karar açısından ifade edilebileceğini öne sürer.

C. THESIS PERMISSION FORM / TEZ İZİN FORMU

(Please fill out this form on computer. Double click on the boxes to fill them)

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Erbaş
Adı / Name : Zeynep
Bölümü / Department : Psikoloji / Psychology

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): The Effect of Criterion Shifts on Familiarity and Recollection in Recognition Memory

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☒ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır.** / Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years**. * ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. * ☐

** Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature ..... **Tarih** / Date .....
(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.