THE ROLES OF FEEDBACK AND PRIOR INFORMATION ON THE EFFECTIVENESS OF THE BASE RATE MANIPULATION FOR RESPONSE BIAS

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

BY

CAVİT DENİZ PALA

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

AUGUST 2019

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

Name, Last name: Cavit Deniz Pala

Signature :

ABSTRACT

THE ROLES OF FEEDBACK AND PRIOR INFORMATION ON THE EFFECTIVENESS OF THE BASE RATE MANIPULATION FOR RESPONSE BIAS

Pala, Cavit Deniz
M.Sc, Department of Psychology
Supervisor: Asst. Prof. Dr. Aslı Kılıç Özhan

August 2019, 82 pages

Recognition memory is the ability to discriminate previously encountered items from new ones. Sensitivity is the degree to which participants discriminate between old and new items, whereas response bias indicates the tendency to classify an item as old. Response bias can be manipulated via methods, such as base rate manipulation, and strength manipulation, etc. The aim of the current study was to investigate the roles of response feedback and prior information in base rate manipulation for criterion shifts. Yes-no word recognition memory tasks were carried out in the study with this purpose. The base rates of test lists and type of information were manipulated in two experiments. The Base rates were 50%, 20% and 20% for three successive tests in Experiment 1A, while the base rates were 50%, 80% and 80% in Experiment 1B. The presence of feedback and prior information was also manipulated across four groups in both experiments. The results indicated that participants made adaptive shifts in their response bias in accordance with the base rates of test cycles when feedback or prior information were provided. However, the criterion did not differ across tests when no supportive cues were presented. Feedback and prior information together caused greater shifts in criteria compared with the groups with the feedback or the prior information alone. Sensitivity also decreased when test lists were comprised of mostly old or mostly new items. The findings were interpreted and compared with the literature.

Keywords: Response Bias, Criterion, Recognition Memory, Feedback

ÖZ

TEPKİ YANLILIĞININ TEMEL ORAN İLE DEĞİŞİMLENMESİNİN ETKİLİLİĞİNDE GERİBİLDİRİM VE ÖN BİLGİNİN ROLLERİ

Pala, Cavit Deniz
M.Sc, Psikoloji Bölümü
Tez Yürütücüsü: Dr. Öğr. Üyesi Aslı Kılıç Özhan

Ağustos 2019, 82 sayfa

Tanıma belleği daha önce karşılaşılan maddeleri yenilerinden ayırt edebilme becerisidir. Duyarlılık eski ve yeni maddelerin ayırt edilebilme düzeyi anlamına gelirken tepki yanlılığı ise maddeleri eski olarak sınıflandırmaya yatkınlığı ifade eder. Tepki yanlılığı, temel oran değişimlemesi, güç değişimlemesi gibi yöntemlerle değişimlenebilir. Bu tezin amacı yanıt geribildiriminin ve ön bilginin temel oran değişimlemesindeki rollerini incelemektir. Bu amaçla evet-hayır tanıma belleği görevleri uygulanmıştır. Test listelerinin temel oranları ve bilgi türü iki deneyde değişimlenmiştir. Deney 1A'da ardışık 3 testin temel oranları %50, %20 ve %20 iken Deney 1B'de temel oranlar %50, %80 ve %80 olarak değişimlenmiştir. Deneylerde ayrıca geribildirim ve ön bilgi 4 grupta değişimlenmiştir. Sonuçlara göre katılımcılar geribildirim ya da ön bilgi sunulduğunda kriterlerini test listelerinin temel oranlarıyla uyumlu şekilde değiştirebilmişlerdir. Ancak, yardımcı ipuçlarının yokluğunda kriter testler boyunca sabit kalmıştır. Geribildirim ve ön bilgi birlikte sunulduğunda kriterde daha büyük oranda değişimlere yol açmışlardır. Ayrıca, test listeleri çoğunlukla eski ya da çoğunlukla yeni maddelerden oluştuğunda duyarlılık düşmüştür. Bulgular yorumlanmış ve alanyazın ile kıyaslanmıştır.

Anahtar Kelimeler: Tepki Yanlılığı, Kriter, Tanıma Belleği, Geribildirim

To Alexandra Elbakyan

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Asst. Prof. Dr. Aslı Kılıç Özhan for her guidance and support since the first day of my graduate studentship. I owe too much to her knowledge, warm attitude, and patience. Her leadership in our laboratory has made it an enjoyable workplace of which I appreciate to be a part.

I would like to thank Assoc. Prof. Dr. Mine Mısırlısoy for being such a great professor to me. It is a great joy to be her student and course assistant.

I would like to thank Assoc. Prof. Dr. Tuğba Uzer Yıldız for accepting to be a member of the examining committee of my thesis.

I owe too much to the people who work in the organizations such as Sci-Hub, Lib-Gen, and Wikipedia. Their efforts and sacrifices for the democratization of information make scientific research meaningful.

I am grateful to my family and friends for being in my life. I will thank them later in person.

I would like to thank my dearest Deniz Korkmaz. There are no words.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Response bias in a recognition memory task is the tendency to classify a test item as old. Participants can make adaptive shifts in their response bias in response to manipulations in the base rates of test lists, memory strength of items, response payoffs, etc. Supportive cues, such as response feedback, prior information regarding the base rates, font cues for conditions, was found necessary to provide shifts in response bias. However, the feasibility of criterion shifts in different conditions and the roles of supportive cues has not been established yet in the literature. The current thesis aimed to provide additional information to the literature regarding the roles of feedback and prior information in base rate manipulations of response bias.

1.1 Recognition Tasks

Recognition memory is the ability to discriminate previously encountered stimuli or events from novel ones. A single item yes-no recognition task (Criss, Aue, & Kılıç, 2014), which is a typical recognition task, consists of study and test phases. In the study phase, participants are expected to memorize a list of stimuli for the following test phase. Words, pictures, nonsense syllables, and numbers are among the mostly presented stimuli in recognition tasks. In the test phase, probe items consisting of a subset of studied items (also called as target or old) along with new (foil or lure) items are presented to participants usually in a mixed order. Participants are asked to distinguish between old and new items by endorsing ("Yes, I studied this item.") or rejecting ("No, I did not study this item.") the item. In most of the recognition tasks, a distractor task follows the study phase in order to eliminate the recency effect or provide a certain study-test lag duration. Simple arithmetical operations, such as number counting tasks or Sudoku puzzles, are among common distractor tasks.

Besides single item yes-no recognition, there are also other variants of recognition tasks which differ in certain aspects, such as stimulus or test type. For instance, confidence judgments (Benjamin, Tullis, & Lee, 2013) in which recognition decisions are made using a Likert type scale is another commonly used test type in recognition tasks. In another test type, known as forced choice test (Malmberg, Criss, Gangwani, & Shiffrin, 2012), multiple items are presented in a test trial, and participants are expected to identify the old item among the new items. In associative recognition tasks (Hockley & Niewiadomski, 2007), on the other hand, memory for associations between stimuli are examined rather than item information. Stimuli in associative recognition tasks are item pairs rather than single items. The test phase of associative recognition tasks consists of target pairs, which are old items studied together, and rearranged (lure) pairs, which are items studied in different pairs. Lastly, some recognition tasks do not have separate study and test parts; rather, participants make recognition judgments for each presented item. Such tasks are called continuous recognition (Estes & Maddox, 1995). These tasks constitute a significant part of recognition memory literature. The main focus of the current thesis is on single item yes-no recognition task since it provides a useful method for response bias research, therefore commonly used in the relevant literature.

1.2 Performance Measurement on Recognition Memory: Signal Detection Theory

In a single item yes-no recognition task, participants' responses constitute a result set of four possibilities (Table 1): A "hit" is endorsement of an old item, while "no" response to an old item causes a "miss". Rejection of a new item is called "correct rejection"; and lastly, incorrect endorsement of a new item is called "false alarm". Hit rate (HR) denotes the proportion of hits to the total number of old items, and false alarm rate (FAR) is the proportion of false alarms to the total number of new items. Performance measurement in a yes-no recognition task is carried out via HRs and FARs.

Table 1

Result Set Matrix in A Single Item Yes-No Recognition Memory Task

	Response	
Item	"Yes"	"No"
Old	Hit	Miss
New	False Alarm	Correct Rejection

Signal detection theory (SDT, Green & Swets, 1966; Macmillan & Creelman, 2005) is a framework by which recognition performance is measured. According to SDT, a subject makes recognition judgments based on the familiarity of the probe item and a subjective decision criterion. Memory strength of a test item indicates how familiar that item is to the subject. Calculation of memory strength is not specified by SDT, but only assumed. Old and new items are assumed to form overlapping distributions on memory strength scale (Figure 1). Old items are generally more familiar, so they have relatively higher values on the scale compared to new items. However, since some new items are more familiar than some of the old items, there is no way to completely distinguish between old and new ones. For this reason, the subject needs to set a threshold by which recognition decision is made: If the memory strength of a test item exceeds that threshold, that item is recognized. This threshold is called criterion. Notice that optimal decision criterion to provide the maximum number of accurate answers is at the intersection point of distributions.

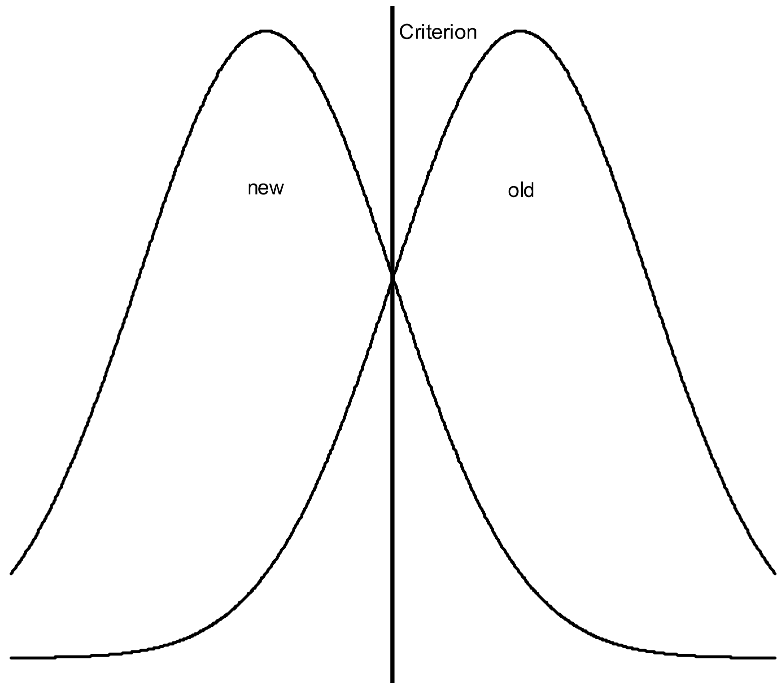


Figure 1. Old and new distributions on memory strength scale according to signal detection theory. The vertical line represents where the criterion is placed on the scale.

1.2.1 Sensitivity and response bias

Considering SDT, one can realize that there are two factors which determine performance: First, the distance between the distributions indicates the extent to which old and new items are discriminated; that is called "sensitivity". Second, criterion placement reveals the tendency of a subject to recognize probe items as old; in other words, "response bias". There are several performance measures for each factor.

The main performance measure for sensitivity is *d'*,

$$d' = z(\mathrm{HR}) - z(\mathrm{FAR})$$

where z denotes the inverse of the normal cumulative distribution of hit and false alarm rates (Macmillan & Creelman, 2005, p. 8). A *d'* value of 0 means no discrimination, and the greater the *d'*, the better the discrimination between old and new items. As for response bias, the main performance measure is *C*, calculated with the following equation (Macmillan & Creelman, 2005, p. 29):

$$C = -0.5(z(\text{HR}) + z(\text{FAR}))$$

A *C* value of 0 signifies a neutral criterion while negative values indicate a tendency towards "liberal" bias, and positive values indicate that the criterion is "conservative". Since the z scores of HR and FAR are derived from normal cumulative distribution, values of 0 and 1 reveal infinite values. In a common method to handle this problem, HR and FAR values of 0 and 1 were replaced by 1 / 2N and 1 - (1 / 2N) values, where N is the total number of trials in the relevant calculation (Macmillan & Creelman, 2005, p. 8). This means that participants are assumed to have at least .5 and at most 1 - .5 hits and false alarms in each calculation.

It is important to note that *d'* and *C* are parametric measures of performance; that is, they rely on the assumptions that familiarity value is normally distributed with equal variance for old and new distributions. However, these assumptions are not always met and further measurement models were developed to account for the data that does not comply with these assumptions (e.g., Wixted, 2007; Yonelinas, 1994). Similarly, nonparametric measures which do not depend on specific assumptions have been developed as alternatives. *A'* is a common nonparametric sensitivity measure,

$$A' = .5 + \frac{(\text{HR}-\text{FAR})(1+\text{HR}-\text{FAR})}{4\text{HR}(1-\text{FAR})} \text{ when HR} \geq \text{FAR}$$

$$A' = .5 - \frac{(\text{FAR}-\text{HR})(1+\text{FAR}-\text{HR})}{4\text{FAR}(1-\text{HR})} \text{ when FAR} \geq \text{HR}$$

and *B''* is often preferred as a nonparametric response bias measure

$$B'' = \frac{\text{HR}(1-\text{HR})-\text{FAR}(1-\text{FAR})}{\text{HR}(1-\text{HR})+\text{FAR}(1-\text{FAR})} \text{ when HR} \geq \text{FAR}$$

$$B'' = \frac{\text{FAR}(1-\text{FAR})-\text{HR}(1-\text{HR})}{\text{FAR}(1-\text{FAR})+\text{HR}(1-\text{HR})} \text{ when FAR} \geq \text{HR}$$

(see Macmillan & Creelman, 1996; Snodgrass & Corwin, 1988 for further discussion.).

1.2.2 The dispute over the calculation of response bias

Whether their assumptions are met or not, parametric measures are frequently used in the literature since they are still considered as reliable, and nonparametric measures are generally equivalents of their parametric counterparts. Yet, the dispute over which performance measures are more reliable is not only limited to the assumptions of measures issue. There are several alternative approaches especially on response bias measurement stemming from different assumptions regarding the decision processes.

Likelihood ratio, β, is another widely used measure of criterion, along with its natural logarithm, ln(β). The rationale behind β is a Bayesian calculation using the probabilities of being old and new. For a certain location on the familiarity scale, the likelihood ratio is its probability of being old divided by the probability of being new. Therefore, β=1 indicates a neutral criterion, greater values signify conservative bias, and lower values signify liberal bias. The simplest expressions for likelihood ratio calculations are as follows (Macmillan & Creelman, 2005, p. 35):

$$\beta = e^{cd'}$$

$$\ln(\beta) = cd' = 0.5(z(\mathrm{FAR})^2 - z(\mathrm{HR})^2$$

(also see Macmillan & Creelman, 2005; Stanislaw & Todorow, 1999 for detailed explications and alternative equations).

In some experiments, it is justifiable to assume that the mean strengths of new distributions for compared conditions are same; so, any differences between FARs reveal a change in criterion. However, for two groups whose FARs are same, *C* is higher for the group with higher HR. That is considered as a misinterpretation among some researchers (Franks & Hicks, 2016; Hicks & Starns, 2014). They suggest that FAR is a better indicator of criterion for these situations. This view is controversial particularly for the experiments with strength manipulation. In these experiments, memory strength of study words is manipulated between conditions via repetition, presentation duration, levels of processing, etc. According to those who use FAR as a criterion measure, new distributions stay constant among conditions. However, differentiation models (Criss, 2006, 2010; Kılıç, Criss, Malmberg, & Shiffrin, 2017; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) suggest that because new words are more easily differentiated among strong words, new distribution for strong lists shifts towards left on the memory strength scale. In this case, a difference in FAR does not necessarily indicate a shift in criterion.

This dispute has been discussed in detail in the following sections. For now, it is crucial to emphasize that the selection of performance measures may change critically the interpretations of experiment results. Moreover, the method through which criterion is manipulated needs to be selected carefully according to the purpose of the research question. In the following section, alternative methods for criterion manipulation have been described together with major findings in the field. This also clarifies the purpose of the current thesis, and its experimental and analysis methods including the research questions, design and performance measures.

1.3 Experimental Methods for Criterion Manipulation

1.3.1 Base rate manipulation

Decision criterion has been a research topic in sensation and perception literature as well as recognition memory since the 1960s (Banks, 1970; Green & Swets, 1966; Healy & Jones, 1975; Swets, Tanner, & Birdsall, 1961). One of the first methods used to manipulate criterion is the base rate manipulation in which the proportion of old and new items are changed. The base rate of a test list is the proportion of old items in a recognition task. When most of the test items are old, the criterion is expected to become more liberal, and vice versa for mostly new lists. Swets et al. (1961) reported the effect of base rate on decision criterion in signal detection task (i.e. In a signal detection task, the base rate is the proportion of target-present trials throughout the task.). In the following years, the effect of base rate on criterion became a research topic for recognition memory as well. Healy and Jones (1975) carried out a recognition memory task with 3 digit numbers as study items. They manipulated the base rates of consecutive test blocks between 50% and 25%

in pseudo-random order. They also informed participants about the base rate of each block at the beginning of that block, gave feedback on whether the response is correct. Participants made confidence judgments on a 4-point Likert scale which results in 3 criteria values. They found a shift only in one of the three criteria between conditions. In the following years, prior information on base rates is commonly used as well as response feedback as supportive cues. Moreover, it has been shown that such cues are necessary, but not always adequate to prompt criterion shifts (Cox & Dobbins, 2011; Estes & Maddox, 1995; Healy & Kubovy, 1977; Koop, Criss, & Malmberg, 2015). In Healy and Jones's (1975) study, it is possible that adaptive shifts in the criterion are hindered by the random order of base rates. In the literature, possible confounding effects of base rate order of consecutive test lists have been mostly overlooked, therefore the results are misinterpreted, as has been discussed in the following sections. In the current thesis, a novel design in terms of the base rate order of test lists has been proposed in order to re-evaluate previous findings in the literature, especially on the role of supportive cues.

Healy and Kubovy (1977) also failed to find criterion shifts between conditions in a similar design. However, they observed a criterion shift in a numerical decision task resulting from base rate change. Observing that sensitivity is higher in the recognition task than numerical decision task, they suggested that high sensitivity cause participants to ignore information about base rates; which has been confirmed repeatedly by later research (Aminoff et al., 2012; Bruno, Higham, & Perfect, 2009; Estes & Maddox, 1995; Hirshman, 1995; Selmeczy & Dobbins, 2013). Later in another study, Healy and Kubovy (1978) observed shifts in criterion in accordance with the base rate in recognition tasks. In recognition tasks, the base rate was manipulated between 50% and 25% in pseudo-random order throughout 8 successive test lists. All participants had prior information about base rates while feedback was manipulated between groups. Participants were able to shift their criterion whether the feedback is presented or not. Note that criterion shifts were observed despite the random order of base rates, contrary to Healy and Jones (1975) and Healy and Kubovy (1977).

Another study investigating the effects of base rate manipulation and feedback revealed conflicting results. Estes and Maddox (1995) carried out experiments with continuous recognition tasks using three types of stimulus: 3-digit numbers, 3-letter syllables, and words. They manipulated the base rate (33% - 67%) along 2 consecutive test lists and feedback (present - absent) in recognition tasks for each stimulus type. In their 2 experiments which base rate manipulation was made within (Experiment 1) or between (Experiment 2) subjects with number and syllable stimuli, criterion shifts were observed in presence of feedback; but there were no changes in other conditions. It is possible that higher sensitivity in word recognition task hindered adaptive changes in criterion as in Healy and Kubovy (1977). For other stimuli, feedback played a critical role to induce shifts in decision criterion.

The essential role of feedback in criterion manipulation was also shown by Rhodes and Jacoby (2007). They discussed conflicting findings in the literature regarding the feasibility of criterion shifts via experimental manipulations, and argued that participants' awareness of the manipulation was likely to be the reason behind the different results. In their study, words were presented on two opposite sides of the screen, and base rates were different on both sides. Participants received feedback for each answer, together with the running count of correct answers which was available throughout the test. In Experiment 1, participants were able to shift their criteria on item-by-item basis in accordance with base rates of the locations. In Experiment 2, they added a cue condition in which half of the participants used different keys to answer the probes on different locations (see Franks & Hicks, 2016; Starns & Olchowski, 2015 for other studies in which response key cues were used). They also assessed whether participants were aware of the manipulation by a post-test questionnaire (Rhodes & Jacoby, 2007). Awareness rose from 25% to 71% when different keys were used. Participants in different keys condition shifted their criteria, and the magnitude of criterion shift increased throughout the test only for the participants who were aware of the manipulation. Participants in the same keys condition failed to shift their decision criteria; besides, awareness had no effect on that condition. Finally, they manipulated the feedback condition in Experiment 3. Half of the participants were given feedback on only the first half of the test, whereas feedback was available for only the second half of the test for the other group. Feedback induced a criterion shift when available, but the shift was not observed when there was no feedback.

A special variant of base rate manipulations is "pure lists" in which all the test list consists of only old (pure-old list) or new (pure-new list) items. *C* cannot be calculated in pure lists, because there are no false alarms in pure-old lists, and no hits in pure-new lists. Therefore, pure lists are compared to standard lists (with 50% base rate) using either HR or FAR. In such designs, a consistent finding is that HR for pure-old and standard lists showed no difference (Cox & Dobbins, 2011; Ley & Long, 1987, 1988; Wallace, 1982; Wallace, Sawyer & Robertson, 1978). Cox and Dobbins (2011) compared pure-old lists and pure-new lists to standard lists and found no criterion differences, or, surprisingly, inconvenient shifts in criterion which is a shift in the opposite direction from what is expected. Some of these experiments (Cox & Dobbins, 2011, Experiment 1; Wallace et al., 1978, Experiment 1) also included prior information about base rates which made no difference in participants' decision processes. A probable explanation for the ineffectiveness of prior information in such designs is that it would make the whole test meaningless if participants benefit from that information. Prior information might lead participants to rely on their mnemonic evidence instead of this external cue. Consistent with this argument, Koop et al. (2015) reported criterion shifts between pure lists and standard lists when feedback is available. Feedback is a relatively implicit cue compared to prior information, and it allows participants to focus on their performance throughout the test. So, participants either realized the composition of test list and act accordingly; or they changed their criterion towards more "reinforced" response.

The base rate manipulation sometimes failed to induce criterion shifts (Cox & Dobbins, 2011; Healy & Jones, 1975; Healy & Kubovy, 1977). However, it was found to be an effective method in some other studies (Estes & Maddox, 1995; Franks & Hicks 2016; Healy & Kubovy, 1978; Koop et al., 2015; Rhodes & Jacoby, 2007). Which designs and supportive cues provide robust shifts are yet to be revealed.

1.3.2 Strength manipulation

Strength manipulation is another method frequently used in recognition memory literature. In such designs, strong words are presented more frequently (Hicks & Starns, 2014; Ratcliff, Clark, & Shiffrin, 1990, Experiment 5-6; Ratcliff, Sheu, & Gronlund, 1992, Experiment 2; Stretch & Wixted, 1998) or for longer durations (Bruno et al., 2009; Hirshman, 1995; Ratcliff et al., 1990, Experiments 1-4; Ratcliff, et al., 1992, Experiment 1, Experiment 3) than weak words so that strong words are encoded better. Study-test delay (short vs. long, Singer & Wixted, 2006), or similarity of new words (similar vs. unrelated new words, Benjamin & Bawa, 2004) are among alternative ways to induce strength manipulation. When strong and weak words are studied and tested in separate lists, HRs are higher and FARs are lower in strong lists than weak lists; that is a major finding in recognition memory literature, called strength based mirror effect (Glanzer & Adams, 1985; Ratcliff et al., 1990; but see Hilford, Glanzer, Kim, & Maloney, 2019). Strength based mirror effect has been considered as a result of criterion shift by some scientists (Hicks & Starns, 2014; Starns, White, & Ratcliff, 2010; Stretch & Wixted, 1998). The rationale is as follows (Figure 2, left panel): New words' familiarity values are not affected by the strength of old words in a test, but strong words are more familiar than weak words on average. So, the optimal criterion is more conservative for strong lists than weak lists. Higher sensitivity and more conservative criterion result in higher HRs and lower FARs. This account relies on the assumption that familiarity of a new word is determined by the pre-experimental probability of it. However, differentiation models (Criss, 2006, 2010; Kılıç et al., 2017; McClelland

& Chappell, 1998; Shiffrin & Steyvers, 1997) suggest a different mechanism for recognition: In a recognition test, a test word is compared to all the words in the study list, its similarity with studied words determines its familiarity value. As the strong words match better with their memory traces, HRs are higher in strong lists. New words in strong lists, likewise, reveal lower familiarity values, because they are less likely to match by chance with well-encoded old words (Figure 2, right panel). Thus, differentiation account does not necessitate criterion shift mechanism to explain strength based mirror effect. Although the dispute has not been resolved, it is known that criterion shift and differentiation both play roles to some extent in recognition tasks in which strength manipulation is involved.

Strong and weak words are presented in the same study list (mixed list) in some experiments. Hirshman (1995) compared *C* values for weak words in mixed lists with pure-weak lists; likewise, *C* values for strong words in mixed lists with pure-strong lists. He found that response bias is determined according to the average strength of study lists (e.g. the criteria for strong words in mixed lists are more liberal than that of pure-strong lists). Whether criterion shifts are possible during a test has been another research question in the literature. In these studies, weak test blocks, which consist of weak words together with new words, and strong test blocks, which consist of strong words and new words, are arranged. Weak and strong blocks are separated from each other via cues such as color (Hicks & Starns, 2014; Stretch & Wixted, 1998) or semantic category (Morrell, Gaitan, & Wixted, 2002; Verde & Rotello, 2007). While some studies did not report criterion shifts during the test (Morrell et al., 2002; Stretch & Wixted, 1998; Verde & Rotello, 2007, Experiments 1-4), others observed such shifts with the help of color cue (Hicks & Starns, 2014) or feedback (Verde & Rotello, 2007). It is still not clear what causes conflicting results, but some differences in the designs of these experiments, such as the order of weak and strong test blocks, the length of blocks (see Hicks & Starns, 2014 for a discussion), or cues could be possible factors. In mixed list designs, differentiation is not possible according to current versions of differentiation models; therefore, it provides a method to investigate the response bias in recognition memory.

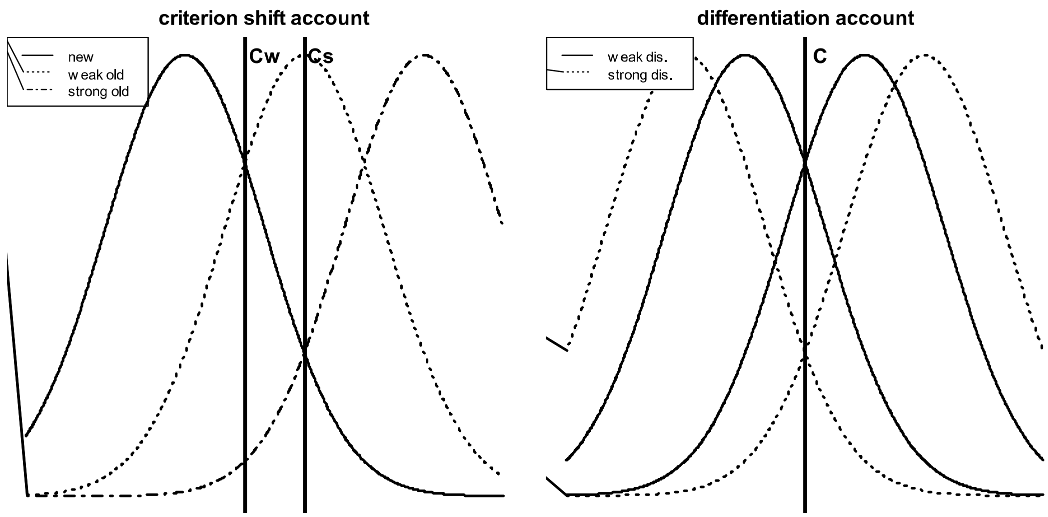


Figure 2. Criterion shift and differentiation accounts for strength based mirror effect. Criterion shift account (left panel) assumes single distribution for new items, and different criteria for strong and weak lists. Differentiation account (right panel) assumes constant criterion and different new distributions for both lists. Cw: Criterion of the weak list. Cs: Criterion of the strong list.

To summarize, strength manipulation between study lists results in a mirror effect possibly stem from criterion shift. However, differentiation models suggest an alternative mechanism which explains mirror effect without criterion shift. There are conflicting results on whether criterion shift is possible in mixed list designs.

1.3.3 Alternative methods

Until now, the literature on base rate and strength manipulations as the most frequently used methods to induce criterion shifts has been summarized. There are also alternative methods in the literature, such as response payoffs (Curran, DeBuse,

& Leynes, 2007; Healy & Kubovy, 1978; Van Zandt, 2000), probabilistic cues (Selmeczy & Dobbins, 2013, 2017), or biased feedback (Han & Dobbins, 2008, 2009). Besides, some stimuli characteristics have been known to affect the decision criterion; such as, word frequency (Glanzer & Adams, 1985; Glanzer, Adams, Iverson, & Kim, 1993) stimulus type, (Scimeca, McDonough, & Gallo, 2011); recognition task type, (Hockley & Niewiadomski, 2007), etc. Some of these studies whose findings are directly related to the current thesis have been described, while others have been excluded since they are out of the scope of current research.

Han and Dobbins (2008) developed a method which relies on biased feedback in order to affect response bias. The biased feedback is similar to the normal item feedback procedure with one exception. On "liberal-biased feedback" condition, only false alarms return "correct" feedback; therefore, participants on this condition never receive negative feedback for their "yes" responses. By contrast, misses return "correct" feedback on "conservative-biased feedback condition. This time, "no" responses are fostered. Han and Dobbins (2008, Experiment 1) divided their participants into two groups, one of whom took three consecutive recognition tasks with liberal, normal and conservative biased feedbacks, successively (LNC condition). For the other group, the test order was reversed (CNL). The experiment revealed that the CNL group were more conservative than the LNC group in first and second tests, while the difference disappeared in the last tests. This indicates a carry-over effect on criterion over successive tests. In Experiment 2, they composed CLL and LCC orders of feedback for two groups. There were no significant differences between groups in the first test, while LCC group were more conservative in the following tests. Thus, two experiments revealed conflicting results. Han and Dobbins continued their research on biased feedback procedure, and found that once the criterion was set as either liberal or conservative, it stayed constant over the following tests with normal feedback (2009). Their participants stated that they were not aware of biased feedback manipulation in post-questionnaire. Their studies were unique in that criterion was induced without participants' awareness. It demonstrates that feedback can influence response bias without arising awareness. Han and Dobbins (2009) named their method as "incremental reinforcement learning", implying that feedback's mechanism of action is differential reinforcement of possible responses.

Selmeczy and Dobbins (2013) adopted a new method to manipulate response bias, called "probabilistic mnemonic cue". Such cues were presented before each test item and gave the probability of that item being old. In their experiment, participants were given "likely old" or likely new" cues before each trial, which was 75% true. They also manipulated the difficulty of the task and the presence of feedback. Participants benefited from the probabilistic cues such that criterion was more liberal for the test words following "likely old" cue. Moreover, the criterion difference between cue conditions was higher in the difficult task rather than the easier task. They reported that the presence of feedback had no effect on criterion or sensitivity in this experiment. They argued that manipulation was explicit to the participants in their study because of the nature of the task, therefore feedback had no additional effects. As for the studies of Estes and Maddox (1995) and Verde and Rotello (2007), feedback caused a criterion shift because it helped participants understand the manipulation. Then, they manipulated the reliability of probabilistic cues (65% - 85%) and observed that the magnitude of criterion shift increased with cue reliability (Selmeczy and Dobbins, 2013, Experiment 2).

1.4 The Effects of Supportive Cues on Response Bias

As summarized, feedback and prior information have been used as auxiliary information in the experiments in which response bias is manipulated. There are various designs which feedback, prior information or other supportive cues (e.g. location, response keys, etc.) are used with different combinations. These studies indicate that the impact of feedback and prior information depends on many factors in experiments, such as stimulus type, difficulty of the task, or presence of other additional cues. For instance, Healy and Jones (1975) and Healy and Kubovy (1977) used prior information and feedback both in their study in which base rate was manipulated, failed to induce consistent changes in response bias. Other studies reported such changes with base rate information whether feedback was present or not (Aminoff et al., 2012; Franks & Hicks, 2016; Frithsen, Kantner, Lopez, & Miller, 2017; Healy & Kubovy, 1978). However, studies with pure list designs surprisingly revealed null results even if prior information was provided to participants (Cox & Dobbins, 2011; Ley & Long, 1987, 1988; Wallace, 1982; Wallace et al., 1978).

As for feedback, it fostered criterion shifts in the absence of prior information in several studies with base rate manipulation (Estes & Maddox, 1995; Koop, et al., 2015; Rhodes & Jacoby, 2007). It also has proved to be beneficial in studies with mixed-list strength manipulation, such that Verde and Rotello (2007, Experiment 5) reported criterion shifts between conditions when feedback is provided, while such changes were not observed in the absence of feedback in many studies (Morell et al., 2002; Starns et al., 2010; Stretch & Wixted, 1998; Verde & Rotello, 2007, Experiments 1-4). However, Hicks and Starns (2014) and Starns et al. (2010) did not observe shifts when feedback was provided to participants.

To summarize, several supportive cues accompany the methods of criterion manipulation. Participants benefit from these cues to make adaptive shifts in their criteria. They have been used in many different ways and combinations, which reveals highly variable results. Further research is required to understand the roles of supportive cues. For this reason, feedback and prior information were manipulated in the current study to further investigate their roles in criterion shifts.

1.5 Feasibility of Multiple Shifts in Response Bias

Many studies in the literature fostered multiple shifts in response bias with alternating liberal and conservative lists. In these designs, participants were expected to adapt their criteria to consecutive changes in conditions. However, multiple shifts often caused participants to fail in adaptive shifts, or previous shifts had lasting effects in response bias throughout the test, therefore caused noise in the results.

Healy and Jones (1975) failed to induce criterion shifts between base rate conditions (25% - 50%) even though they used prior information and feedback. They changed the base rates of successive test blocks consisting of 40 items in such order that there were no three consecutive blocks of the same condition. There were 8 test blocks in the test, so, participants were expected to make several shifts along the test. In the following years, Healy and Kubovy (1977, 1978) employed very similar designs in terms of the base rate order, failed to observe criterion shifts in one of the experiments (1977), and reported such shifts in the other (1978). Estes and Maddox (1995, Experiment 1) conducted an experiment with base rate manipulation between two consecutive test blocks and observed criterion shifts with syllables and digit stimuli when feedback is available, but criterion was not different with words or in the absence of feedback. Note that null results in this experiment could be related to other factors than base rate order because they replicated the same results in their Experiment 2 in which base rate was manipulated between subjects. In the following years, Aminoff et al. (2012) and Franks and Hicks (2016) managed to induce criterion shifts between test blocks consisting of 6 to 9 items with the help of color cues and prior information. Additionally, Heit, Brockdorff, and Lamberts (2003) and Rhodes and Jacoby (2007) observed item-by-item shifts in line with base rates. These findings indicate that within-list criterion shifts are possible under certain circumstances which are not clearly established yet.

Feasibility of multiple shifts was also tested via studies with strength manipulation. Verde and Rotello (2007) divided recognition tasks into 80-item strong and weak test blocks, and compared the groups who started with strong block to the group who started with weak block. They found that response bias throughout the test was determined by the initial block. Later, Hicks and Starns (2014) discovered that when the test blocks were long enough, response bias throughout the test was different between the groups whose first blocks were strong or weak (Hicks & Starns, 2014; Experiments 3-4, 40-item condition; Verde & Rotello, 2007). However, there were no differences in response bias resulting from the first block's strength (Hicks & Starns, 2014; Experiments 3-4, 20-item and 10-item conditions; Starns et al., 2010). These experiments have shown that the first shift had lasting effects throughout the tests, which would prevent observing adaptive shifts in criterion across strength or base rate conditions. The experiments of Han and Dobbins (2008, 2009) revealed parallel results to the studies with base rate or strength manipulation. They employed biased feedback method to induce criterion shifts. An important aspect of their study was the order of feedback conditions: It started with either conservative or liberal in all experiments; therefore, the extent to which feedback affected their initial criterion was unknown. In the following tests, the feedback condition has changed to either opposite or normal, while participants' criterion had already been affected by the first test. Carry-over effect might have hindered probable adaptive changes. This was an important finding; however, the magnitude of the initial criterion shift is still unknown, besides, the magnitude of the following shifts is unreliable because of noisy designs. Therefore, Han and Dobbins' experiments share the same shortcomings as the studies with base rate manipulations.

To conclude, the studies which fostered multiple shifts in response bias revealed that previous shifts in response bias affect subsequent shifts and confound the results. These studies did not observe the initial criteria of the subjects; thus the magnitude of initial shifts remained unknown, so did the effects of the initial shifts. In the current study, the experiments began with a standard test in terms of the base rate, then subsequent tests were comprised of either mostly-old or mostly-new words. Thus, it was possible to observe participants' initial criteria and the degree to which they made adaptive shifts.

1.6 Current Study

The methods of criterion manipulation and some major findings in the literature have been summarized until this section. As described, the variability of the designs in the literature is remarkable, as well as conflicting results. The first objective of the current thesis is to propose a design that controls for the variations stemming from the order in which criterion is manipulated along successive test lists. The second objective is to observe how prior information and feedback affect the magnitude (or existence) of criterion shifts arising from the base rate manipulation. The current study aims to make a contribution for understanding the role these supportive cues play in base rate manipulations and interpret the conflicting results in the literature.

In the current study, a novel design was proposed regarding the order in which base rate is manipulated across test lists. Base rate manipulation was selected as the method of manipulation, because it does not affect the composition of study lists, therefore the results cannot be attributed to other factors as in experiments with strength manipulation. In two experiments, participants took single item yes-no recognition memory tasks with Turkish words as stimuli. Each experiment had three consecutive study-test cycles with varying base rates. In Experiment 1A, the base rates of the test lists were 50%, 20%, and 20%, successively. Participants began the task with a standard test list so as to set their initial criteria without any manipulation. After their default criteria were set and observed with the first test, they continued with the following two tests with 20% base rates. Thus, the extent to which participants made adaptive shifts in their decision criteria, which was neutral-to-conservative shifts, was observed. The only difference in Experiment 1B was that the base rates of test lists were 50%, 80%, and 80% in order to observe neutral-to-liberal shifts. The presence of prior information and feedback was also manipulated in the experiments. In both experiments, participants were divided into 4 groups in which they were presented prior information and feedback both, only feedback, only prior information, or no information of either kind.

d' and *C* was calculated as the SDT measures of sensitivity and response bias. These are the most common measures used in the literature; besides, since the composition of the study lists was the same in terms of encoding, *d'* and *C* provides legitimate measures for performance.

This is the first study in which prior information and feedback were orthogonally manipulated in a design with base rate manipulation. In this way, it was possible to observe their effects in the same experiment. This is especially important to understand the way feedback affects criterion. Turner, Van Zandt, and Brown (2011) suggested that feedback provides accumulating information about old and new distributions, therefore help participants adapt their response bias accordingly. Some researchers discussed the effect of feedback in terms of awareness (Hicks & Starns, 2014; Koop et al., 2015; Rhodes & Jacoby, 2007; Selmeczy & Dobbins, 2013). Hicks and Starns (2014) stated that feedback did not affect the size of criterion shifts induced by color cues. Selmeczy and Dobbins (2013) likewise reported no additional effects of feedback on response bias, while bias was affected by the different probabilistic cues. These results indicate that when participants were made aware of the difference between conditions, the presence or absence of feedback made no difference in response bias. By contrast, Han and Dobbins (2008, 2009) induced criterion shifts with biased feedback when participants were not aware of the manipulation. Biased feedback mechanism works without awareness, probably through differential reinforcement of responses. This view shares the same approach to criterion adjustment mechanism as the models which assumes the decision criterion is adjusted on trial-by-trial basis according to positive and negative outcomes of responses (Maddox, 2002; Treisman & Williams, 1984). If that is the case, feedback should affect response bias whether participants are provided with prior information or not. In Experiment 1A, a post-test questionnaire was administered to the participants so as to observe whether they were aware of the base rates. Besides, the systematic manipulation of the cues allowed to observe the additional effect of feedback when prior information was provided to participants. The effect of feedback was interpreted with the help of these information.

The initial criterion was expected to be kept constant throughout the test cycles in the absence of feedback or prior information. Supportive cues were found to be necessary to induce shifts in criterion in the literature. However, in experiment 1A, response bias was expected to be more conservative in Test 2 and Test 3 in the groups which feedback or prior information was provided. In experiment 1B, on the other hand, the criteria in Test 2 and Test 3 were expected to be more liberal than Test 1 in these groups. Prior information was expected to be utilized to the same degree throughout the tests, while the utilization of feedback would take some time until a certain amount of response feedback accumulates. This might cause the response bias to differ from Test 2 to Test 3 via greater adaptive shifts in the groups with response feedback. Sensitivity was expected to be the same across test cycles in all groups, except for the possible decline caused by vigilance, and output interference (Criss, Malmberg, & Shiffrin, 2011) within the test sessions. Lastly, the magnitude of criterion shifts was expected to be greater when feedback was provided in addition to prior information.

CHAPTER 2

METHOD

The current thesis was registered in Open Science Framework. All materials are available at osf.io/b23a8/. The experiments conducted were approved by the METU Human Subjects Ethics Committee (Appendix A).

2.1 Participants

A hundred and twenty students from Middle East Technical University participated in Experiment 1A, an additional 124 students participated in Experiment 1B for extra course credit. There were 87 male and 157 female participants with the average age of 21.86 (SD = 3.93). Participants who had poor performance (i.e. $d' < .5$ along the task, or $d' < .1$ in a test cycle.) or outlier C values (i.e. C is 3 standard deviations higher or lower than the group average of the relevant test cycle) were excluded from the analysis. In Experiment 1A, 5 participants' data were excluded because of poor performance, leaving 115 participants for the subsequent analysis. In Experiment 1B, 4 participants were excluded because of poor performance and 1 participant due to an outlier C score, leaving 119 participants for the analysis.

2.2 Stimulus

Experiments 1A and 1B were word recognition tasks. Experiment 1A consisted of 300 study words, and additional 240 new test words, while Experiment 2A consisted of 300 study words and 60 new test words. The words were randomly selected from Turkish Word Norms (Tekcan & Göz, 2005) for each participant.

2.3 Design

Experiment 1A (Table 2) consisted of 3 study-test cycles with distractor summation tasks separating each study and test session. Participants studied 100 words and were tested with 100 test words including old and new words in each cycle. The base rates of 3 test lists were 50%, 20%, and 20% in succession. The type of information was also manipulated in the experiment as a between subject factor with 4 conditions: No information (No Inf), prior information (PI) only, feedback (FB) only, and feedback and prior information (FB+PI) both. In No Inf group, participants were neither provided with any information about the base rates of test lists nor their response accuracy for any items. PI only group was informed about the base rate before each test session begins, and the base rate of the test session was presented on the screen during the tests. FB only group was provided with feedback after each answer. Finally, FB+PI group received both prior information and feedback. After the recognition task, participants were given a post-test questionnaire to assess whether they were aware of the base rates; and if they were, whether they used some strategies based on that information during tests.

Experiment 1B (Table 2) were the same as Experiment 1A in all aspects except for two differences: The base rates of test lists were 50%, 80%, and 80% in succession, and there was no post-test questionnaire.

2.4 Procedure

The experiments were conducted by a software developed on MATLAB. Participants were assigned to conditions by computer according to their participation order. After signing the informed consent (Appendix B), participants were invited to the laboratory room in which the experiment was carried out, then they were given the instructions of the experiment written on the computer screen (Appendix C). The experiment then started with the first study session. In study sessions, 100 words were presented successively in white font color, written in 65-size Arial font in the center of the screen with a dark grey background. Each word was presented for 1 second with a 100 ms interstimulus interval. The durations were determined arbitrarily to form a list difficult enough to produce criterion shifts according to the literature (Koop et al. 2015; Rhodes & Jacoby, 2007; Selmeczy & Dobbins, 2013). A 45-second summation task as a distractor followed the study list

(Criss et al., 2014; Koop et al., 2015), then the test part began. Participants in PI only and FB+PI groups were shown the base rate of the following test at the beginning of the test session (Appendix D). Test words were presented successively with the question asking whether they have seen the word in the study list, written on the top of the screen with smaller font size (Appendix E). Base rate information, if given, was written on the top right corner of the screen in yellow during the test. Participants gave their answers via keyboard by pressing "z" key for "yes" response, and "m" key for "no" response. Responses were recorded by the software. The experiments took 20 minutes on average. Participants in Experiment 1A were administered the post-test questionnaire verbally by the experimenter following the recognition task (Appendix F).

Table 2

The Base Rates in Experimental Conditions of Experiment 1A and 1B

	Test Cycle			
	Type of Information	Test 1	Test 2	Test 3
Experiment 1A	No Inf	50%	20%	20%
	PI only	50%	20%	20%
	FB only	50%	20%	20%
	FB+PI	50%	20%	20%
Experiment 1B	No Inf	50%	80%	80%
	PI only	50%	80%	80%
	FB only	50%	80%	80%
	FB+PI	50%	80%	80%

CHAPTER 3

RESULTS

3.1 Data Transformation and Exclusion

HRs and FARs were calculated for each condition and corrected if necessary (see Table 3 and Table 4 for the descriptive results). Specifically, HR and FAR values of 0 and 1 were adjusted to l/(2N) and 1 - 1/(2N) according to Macmillan and Creelman (2005, p. 8). Signal detection response bias measure *C* and sensitivity measure *d'* were calculated by using hit and false alarm rates of the participants. *d'* and *C* values were compared with 4 x 3 mixed ANOVA analyses via R, JASP, and SPSS software. The assumptions of normality, homogeneity of variances, and sphericity were controlled by the tests of Kolmogorov-Smirnov, Levene's, and Mauchly's. The test results were reported and necessary adjustments in test statistics were made in case of any violations. Note that although the assumptions of normality and homogeneity of variances were violated in few cases, ANOVA analyses were carried out without any data transformations, relying on the robustness of the test. ΔC values denoting the magnitude of criterion shifts between base rate conditions were also calculated. ΔC scores of PI only, FB only and PI only conditions were compared via planned pairwise comparisons after the normality of the distributions were checked. All post-hoc comparisons were carried out using Bonferroni adjustment for *p* values except for the planned contrasts which have been mentioned.

3.2 Experiment 1A

3.2.1 Sensitivity

The sensitivity scores of *d'* which were significantly non-normal according to Kolmogorov-Smirnov tests of normality are as follows: No Inf group Test 1

($D(30) = .166$, $p = .034$); PI only group Test 1 ($D(28) = .168$, $p = .043$), Test 2 ($D(28) = .187$, $p = .013$), and Test 3 ($D(28) = .205$, $p = .004$).

The sensitivity was analysed via *d'* values with a 4 x 3 mixed-factor ANOVA (see Table 5 for the descriptive results). Type of information (No Inf, PI only, FB only, FB+PI) was the between subject factor, while test cycle (1, 2, 3) was the repeated measure factor of the experiment. Main effect of test cycle was significant ($F(2, 222) = 8.559$, $MSE = 1.255$, $p < .001$, $\eta_p^2 = .072$), while type of information main effect was not significant ($F(3, 111) = 0.387$, $MSE = 0.244$, $p = .762$, $\eta_p^2 = .010$). There was not significant interaction between conditions ($F(6, 222) = 0.599$, $MSE = 0.088$, $p = .731$, $\eta_p^2 = .016$). Pairwise comparisons for test cycle (Figure 3) revealed that sensitivity decreased from Test 1 ($M = 1.408$, $SD = 0.481$) to Test 2 ($M = 1.265$, $SD = 0.568$, $t(114) = 2.898$, $p = .014$, $d = 0.270$) and Test 3 ($M = 1.204$, $SD = 0.597$, $t(114) = 4.303$, $p < .001$, $d = 0.401$). However, sensitivity was not different between Test 2 and Test 3 ($t(114) = 1.129$, $p = .784$, $d = 0.105$).

To further investigate the changes in sensitivity throughout test cycles, test lists were divided into 4 blocks with 25 items. HRs and FARs were calculated for each block, and blocks with 0 targets or foils were excluded. 2 blocks in individual data were excluded because they had 0 targets. *d'* values were calculated based on HR and FAR values of the blocks. The test blocks were numbered according to their order in the whole task. The pattern of sensitivity throughout the task has been depicted in Figure 4.

A' values were also calculated as a nonparametric alternative measure of sensitivity (Table 6). The same analysis with mixed ANOVA design was calculated as in *d'* values. The data demonstrated the same pattern such that all omnibus and post-hoc analyses revealed the same results in terms of significance (Figure 5). To save space, only the descriptive tables and figures were included in the results.

Table 3

Mean and Standard Deviation Values for HR and FAR in Experiment 1A

Test Cycle	Type of Inf.	HR		FAR		N
Test 1	FB+PI	0.710	(0.119)	0.248	(0.100)	30
	FBonly	0.725	(0.079)	0.245	(0.100)	27
	NoInf	0.673	(0.127)	0.177	(0.106)	30
	PIonly	0.664	(0.107)	0.174	(0.104)	28
Test 2	FB+PI	0.527	(0.185)	0.133	(0.072)	30
	FBonly	0.596	(0.170)	0.186	(0.106)	27
	NoInf	0.640	(0.171)	0.195	(0.119)	30
	PIonly	0.475	(0.184)	0.114	(0.068)	28
Test 3	FB+PI	0.450	(0.144)	0.108	(0.049)	30
	FBonly	0.530	(0.175)	0.139	(0.082)	27
	NoInf	0.600	(0.168)	0.191	(0.132)	30
	PIonly	0.429	(0.210)	0.104	(0.082)	28

Table 4

Mean and Standard Deviation Values for HR and FAR in Experiment 1B

Test Cycle	Type of Inf.	HR		FAR		N
Test 1	FB+PI	0.696	(0.143)	0.313	(0.127)	30
	FBonly	0.733	(0.105)	0.218	(0.119)	30
	NoInf	0.635	(0.125)	0.211	(0.137)	30
	PIonly	0.705	(0.103)	0.239	(0.105)	29
Test 2	FB+PI	0.810	(0.108)	0.473	(0.167)	30
	FBonly	0.795	(0.065)	0.403	(0.170)	30
	NoInf	0.583	(0.135)	0.213	(0.177)	30
	PIonly	0.763	(0.092)	0.378	(0.155)	29
Test 3	FB+PI	0.838	(0.109)	0.532	(0.221)	30
	FBonly	0.842	(0.068)	0.458	(0.174)	30
	NoInf	0.563	(0.168)	0.229	(0.153)	30
	PIonly	0.764	(0.103)	0.372	(0.199)	29

Table 5
d' Values in Experiment 1A

Test Cycle	Type of Inf.	Mean	SD	N
Test 1	FB+PI	1.300	0.475	30
	FBonly	1.356	0.430	27
	NoInf	1.484	0.460	30
	PIonly	1.492	0.548	28
Test 2	FB+PI	1.246	0.543	30
	FBonly	1.255	0.694	27
	NoInf	1.347	0.545	30
	PIonly	1.206	0.501	28
Test 3	FB+PI	1.147	0.530	30
	FBonly	1.230	0.642	27
	NoInf	1.241	0.539	30
	PIonly	1.201	0.699	28

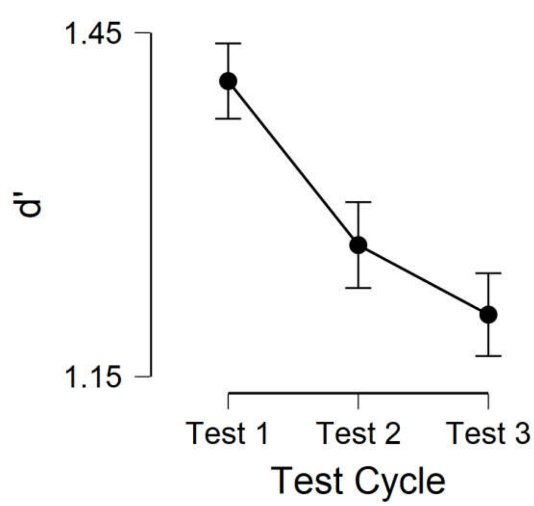


Figure 3. d' values as a function of test cycle in Experiment 1A. Error bars represent standard errors.

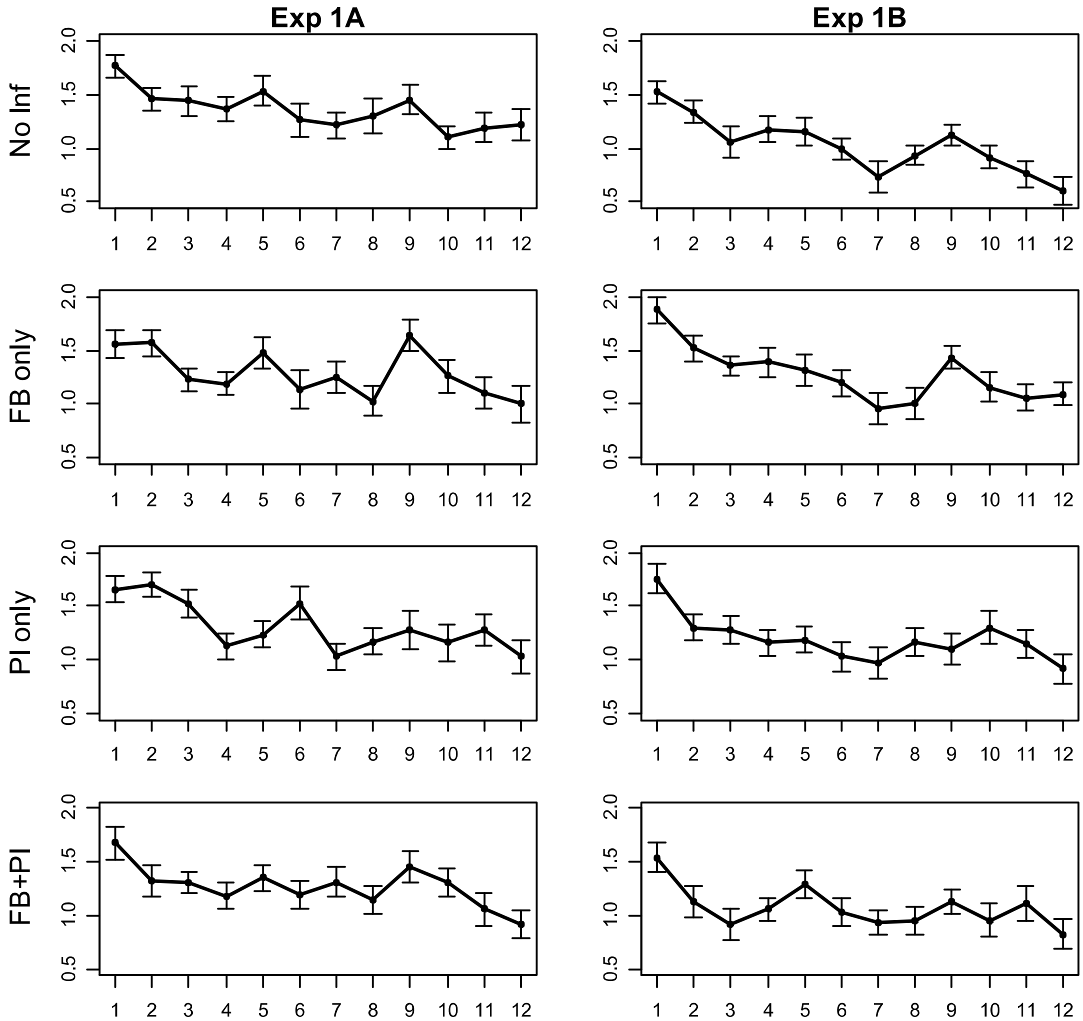


Figure 4. d' values as a function of test block in Experiments 1A and 1B. Error bars represent standard errors.

Sensitivity decreased in Test 2 compared to Test 1, when the base rates of test lists was changed. However, there were no significant differences between Test 2 and Test 3. Therefore, the likely reason behind the decline in sensitivity is the introduction of mostly-new test lists. There were no significant differences between type of information groups in terms of sensitivity.

Table 6
A' Values in Experiment 1A

Test Cycle	Type of Inf.	Mean	SD	N
Test 1	FB+PI	0.813	0.068	30
	FBonly	0.823	0.056	27
	NoInf	0.835	0.057	30
	PIonly	0.833	0.053	28
Test 2	FB+PI	0.794	0.083	30
	FBonly	0.790	0.096	27
	NoInf	0.811	0.075	30
	PIonly	0.788	0.070	28
Test 3	FB+PI	0.774	0.086	30
	FBonly	0.788	0.091	27
	NoInf	0.795	0.081	30
	PIonly	0.774	0.088	28

3.2.2 Response bias

Tests of normality for *C* scores showed that the assumption of normality is violated in FB only group, Test 1 ($D(27) = .195$) Besides, variances of the group means were heterogeneous in Test 3, according to the Levene's test ($F(3,111) = 3.637$, $p = .015$).

C scores were analysed with a 4 (type of information) x 3 (test cycle) mixed-factor ANOVA with repeated measures on the test cycle (see Table 7 for the descriptive results). Analysis revealed a significant shift in response bias across the test cycles as a result of base rate manipulation ($F(2, 222) = 202.352$, $MSE = 5.433$, $p <.001$, $\eta_p^2 = .480$). There was also the main effect of type of information ($F(3, 111) = 7.345$, $MSE = 1.626$, $p < .001$, $\eta_p^2 = .166$). Moreover, significant interaction between test cycle and type of information ($F(6, 222) = 9.391$, MSE = 0.498, $p <.001$, $\eta_p^2 = .202$) indicated that type of information affected the magnitude of response bias shifts. To interpret the differences between conditions, pairwise comparisons of the test cycles for each type of information were carried out. According to pairwise comparisons of test cycles for each group (Figure 6), No Inf group did not benefit from base rate information, such that their criteria did not differ significantly from Test 1 to Test 2 ($t(29)= 0.180$, $p = 1$, $d = 0.033$), from Test 1 to Test 3 ($t(29)= 1.397$, $p = .498$, $d = 0.255$), or from Test 2 to Test 3 ($t(29)= 1.426$, $p = .477$, $d = 0.260$). In PI only group, criterion was more conservative compared to Test 1 in Test 2 ($t(27)= 5.968$, $p < .001$, $d = 1.128$) and in Test 3 ($t(27)= 7.738$, $p < .001$, $d = 1.462$). However, it did not differ significantly between Test 2 and Test 3 ($t(27)= 2.268$, $p = .076$, $d = 0.429$). As for FB only group, criterion increased from Test 1 to Test 2 ($t(26)= 4.422$, $p < .001$, $d = 0.851$), from Test 1 to Test 3 ($t(26)= 7.091$, $p < .001$, $d = 1.365$), and from Test 2 to Test 3 ($t(26)= 3.246$, $p = .005$, $d = 0.625$). In FB+PI group, likewise, response bias became more conservative from Test 1 to Test 2 ($t(29)= 7.934$, $p < .001$, $d = 1.449$), from Test 1 to Test 3 ($t(29)= 10.270$, $p < .001$, $d = 1.875$), and from Test 2 to Test 3 ($t(29)= 3.019$, $p = .001$, $d = 0.551$).

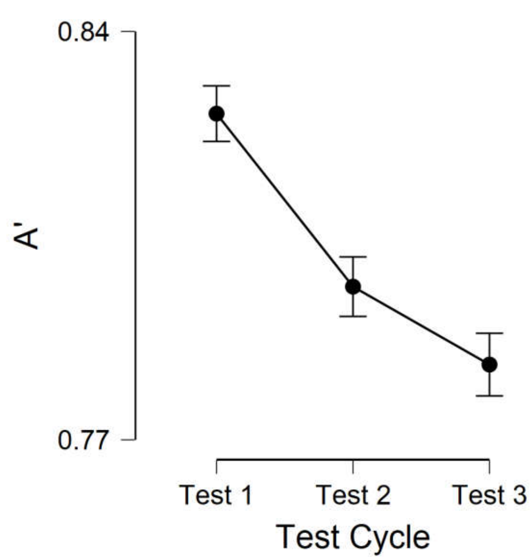


Figure 5. A' values as a function of test cycle in Experiment 1A. Error bars represent standard errors.

To investigate whether the extent of criterion shifts was affected by the type of information, the magnitude of criterion shifts of PI only, FB only and FB+PI groups were compared via contrast analyses. First, Δ*C* values were calculated which denotes the *C* difference between Test 1 and average of Test 2 and Test 3 [(C_{Test3} + C_{Test2}) / 2 – C_{Test1}] (Table 8). Then, contrast analyses were run for each pair of three groups. Analysis indicated that FB+PI group made significantly greater adaptive shifts than FB only group ($t(82)=2.276$, $p=.025$, $d = 0.625$), but the magnitude of criterion shift was not significantly higher than that of PI only group ($t(82)=1.524$, $p=.131$, $d = 0.412$). PI only and FB only groups also did not differ from each other ($t(82)=0.754$, $p=.453$, $d = 0.192$).

Table 7
C Values in Experiment 1A

test cycle	Type of Inf.	Mean	SD	N
test 1	FB+PI	0.064	0.241	30
	FBonly	0.064	0.244	27
	NoInf	0.262	0.330	30
	PIonly	0.301	0.320	28
test 2	FB+PI	0.548	0.334	30
	FBonly	0.346	0.320	27
	NoInf	0.273	0.403	30
	PIonly	0.677	0.363	28
test 3	FB+PI	0.711	0.207	30
	FBonly	0.531	0.281	27
	NoInf	0.350	0.398	30
	PIonly	0.804	0.435	28

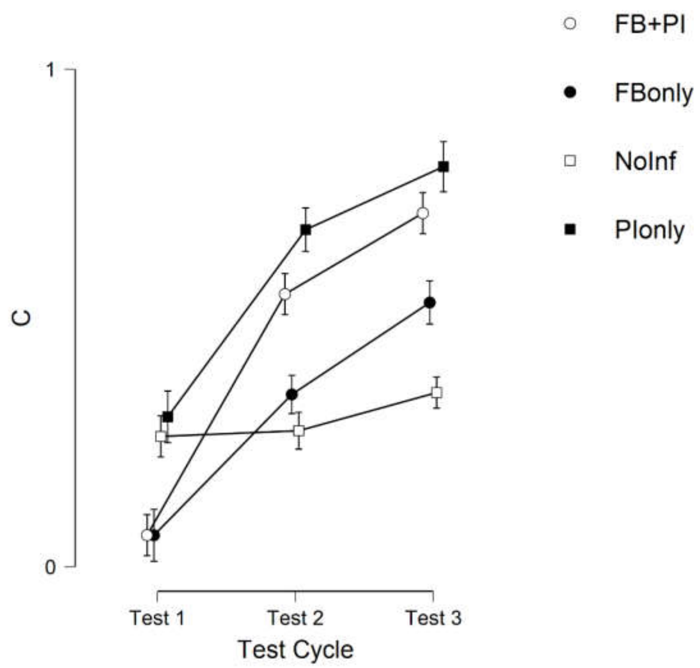


Figure 6. *C* values as a function of test cycle in Experiment 1A. FB: feedback, PI: prior information, NoInf: no information. Error bars represent standard errors.

Table 8
ΔC Values in Experiment 1A

Type of Inf.	Mean	SD	N
FB+PI	0.566	0.277	30
FBonly	0.375	0.333	27
PIonly	0.439	0.336	28

Shifts in response bias were also examined as a function of test block, as in sensitivity analysis. *C* scores were calculated based on the HR and FAR values of blocks (Figure 7).

Response bias was also analysed via *B''* values as a nonparametric alternative (Table 9). The analysis revealed the same pattern as *C* analysis (Figure

8). Lastly, to understand whether criterion reached to its optimal value, likelihood ratio β values for each condition were calculated (Table 10). Note that the optimal criterion for standard tests is $\beta = 1$, while it is 4 for the tests with 25% base rate. As seen in the Table 10, the shifts were suboptimal even if there were substantial shifts.

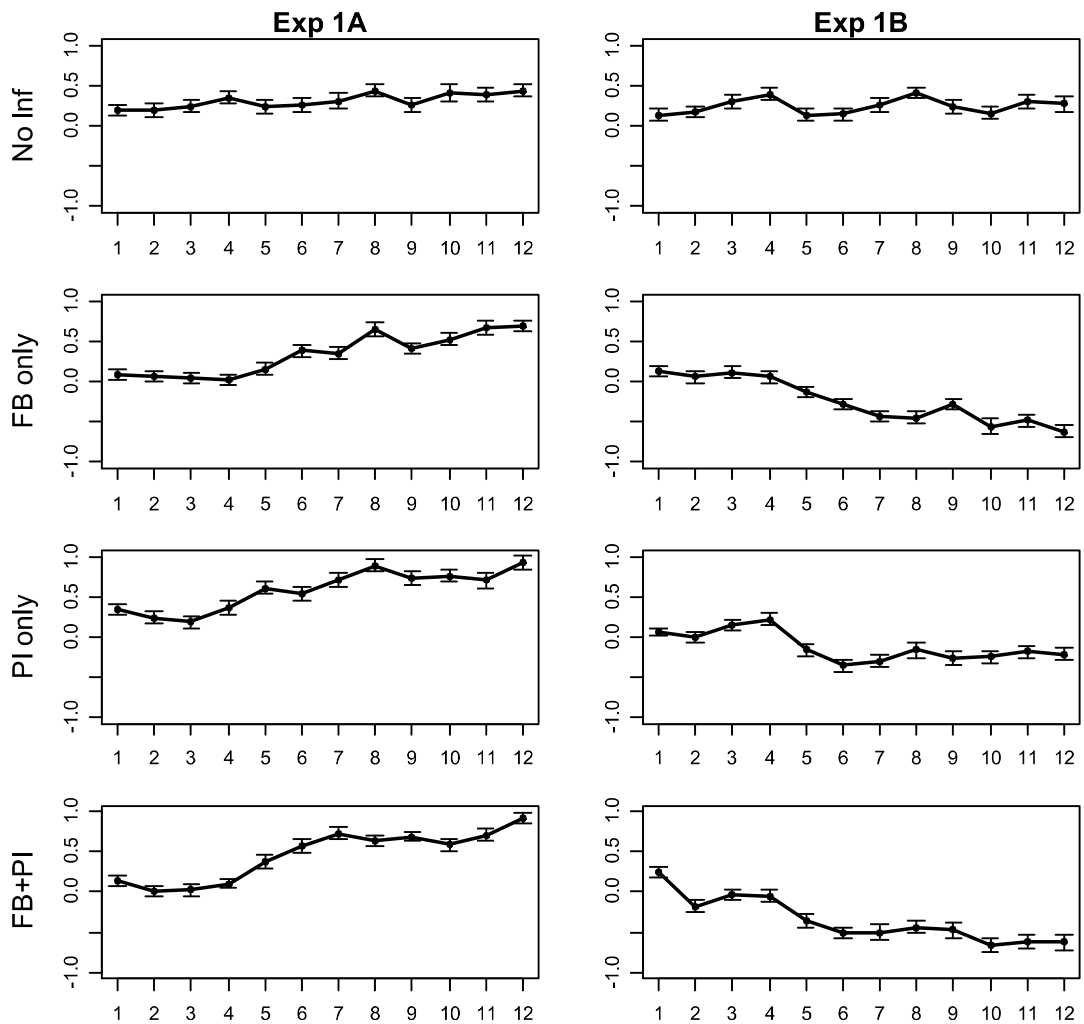


Figure 7. C values as a function of test block in Experiments 1A and 1B. Error bars represent standard errors.

The base rate manipulation caused substantial shifts in criteria when prior information or feedback is provided to participants, but there were no shifts in the absence of these supportive cues. In the groups with response feedback, response bias was more conservative in Test 3 than Test 2. However, response bias did not differ significantly between Test 2 and Test 3 in PI only group. Participants in the FB+PI group made greater shifts than FB only group, but the magnitude of shifts was not significantly greater than PI only group.

Table 9
B'' Values in Experiment 1A

test cycle	Type of Inf.	Mean	SD	N
test 1	FB+PI	0.043	0.171	30
	FBonly	0.066	0.223	27
	NoInf	0.223	0.288	30
	PIonly	0.264	0.286	28
test 2	FB+PI	0.345	0.198	30
	FBonly	0.222	0.261	27
	NoInf	0.183	0.275	30
	PIonly	0.412	0.195	28
test 3	FB+PI	0.428	0.184	30
	FBonly	0.343	0.163	27
	NoInf	0.247	0.258	30
	PIonly	0.461	0.244	28

3.2.3 Post-test questionnaire

Participants' awareness of the base rates of test lists has been determined by the experimenter via post-test questionnaire. Participants who noticed the most of the test items were old in Test 2 and Test 3 but not in Test 1 were classified as aware.

Table 10
β *Values in Experiment 1A*

test cycle	Type of Inf.	Mean	SD	N
test 1	FB+PI	1.114	0.306	30
	FBonly	1.337	1.252	27
	NoInf	1.734	1.019	30
	PIonly	2.235	2.438	28
test 2	FB+PI	1.998	0.846	30
	FBonly	1.737	1.175	27
	NoInf	2.093	3.783	30
	PIonly	2.367	1.286	28
test 3	FB+PI	2.393	1.177	30
	FBonly	1.897	0.579	27
	NoInf	1.726	0.824	30
	PIonly	3.039	2.220	28

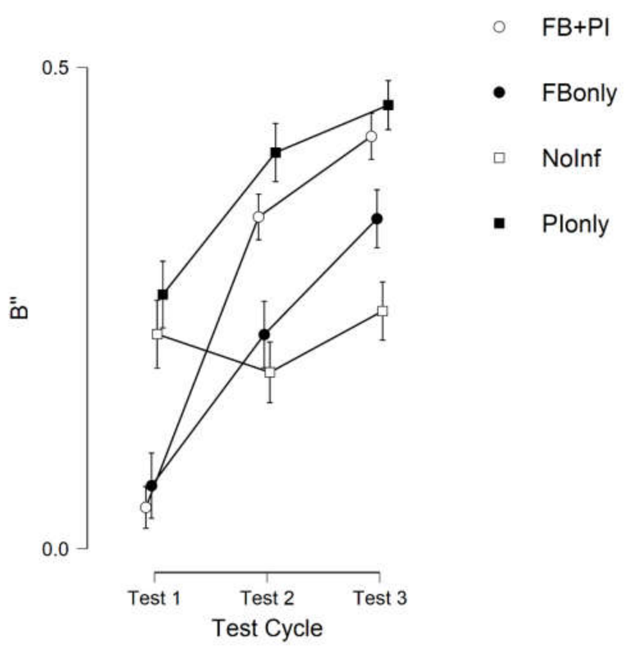


Figure 8. B'' values as a function of test cycle in Experiment 1A. FB: feedback, PI: prior information, NoInf: no information. Error bars represent standard errors.

In No Inf group, only 2 of the 30 participants told that they noticed the number of old items was fewer in Test 2 and Test 3 compared with Test 1. However, almost all participants to whom prior information was presented remembered the base rates. In PI only group, 25 participants remembered the base rates while 3 could not. Similarly, 28 out of 30 participants remembered the base rates while 2 participants had no clear answers in FB+PI group. In FB only group, on the other hand, 8 participants correctly guessed the base rates while 17 could not, and 2 participants did not give clear responses. To investigate the effect of awareness of the participants in FB only group, ΔC values of aware and unaware participants were compared. Aware (M = 0.476, SD = 0.354) and unaware (M = 0.353, SD = 0.337) participants both shifted their criteria, and the t-test analysis comparing the groups did not indicate significant differences in the degree of change ($t(23) = 0.836$, p = .412, d = .358).

Feedback information alone was adequate to induce shifts whether participants were aware of the base rate manipulation or not.

3.3 Experiment 1B

3.3.1 Sensitivity

According to Kolmogorov-Smirnov tests of normality, FB+PI group's Test 3 scores of *d'* was not normally distributed ($D(30) = .163$, $p = .041$).

A 4 (type of information: No Inf, PI only, FB only, FB+PI) x 3 (test cycle: 1, 2, 3) mixed-factor ANOVA was conducted to analyse *d'* scores (Table 11). The analysis revealed that sensitivity was affected by test cycle ($F(2, 230) = 9.808$, $MSE = 1.481$, $p < .001$, $\eta_p^2 = .079$), but the main effect of type of information was not significant ($F(3, 115) = 1.863$, $MSE = 0.920$, $p = .140$, $\eta_p^2 = .046$), neither was the interaction ($F(6, 230) = 1.321$, $MSE = 0.200$, $p = .248$, $\eta_p^2 = .033$). As in Experiment 1A, Post-hoc comparisons for test cycle (Figure 9) indicated that *d'* was higher in Test 1 (M = 1.290, SD = 0.519) than Test 2 (M = 1.113, SD = 0.534, $t(118) = 3.431$, $p = .002$, d = .315) and Test 3 (M = 1.085, SD = 0.504, $t(118) = 4.181$, $p < .001$, d =

.383); however, there was not significant difference between Test 2 and Test 3 in terms of *d'* ($t(118) = 0.563$, $p = 1$, $d = .052$).

Sensitivity as a function of test blocks was also examined in Experiment 1B with the same method as in Experiment 1A (Figure 4). *A'* scores were calculated for Experiment 1B (Table 12). Similar patterns were observed the analyses with *A'* and *d'* values (Figure 10).

Sensitivity decreased from Test 1 to Test 2 when base rate of test lists was changed. However, there was not a significant decline in Test 3 compared to Test 2. The results were parallel to the results of Experiment 1A.

Table 11
d' Values in Experiment 1B

Test Cycle	Type of Inf.	Mean	SD	N
Test 1	FB+PI	1.078	0.561	30
	FBonly	1.502	0.433	30
	NoInf	1.256	0.519	30
	PIonly	1.328	0.486	29
Test 2	FB+PI	1.018	0.439	30
	FBonly	1.138	0.622	30
	NoInf	1.192	0.610	30
	PIonly	1.105	0.449	29
Test 3	FB+PI	0.988	0.529	30
	FBonly	1.171	0.462	30
	NoInf	1.032	0.449	30
	PIonly	1.149	0.571	29

3.3.2 Response bias

Non-normally distributed *C* scores for PI only group are as follows: Test 1 ($D(29) = .169$, $p = .034$), Test 2 ($D(29) = .175$, $p = .024$). Besides, the variances of the *C* scores for the groups were heterogeneous in Test 2 ($F(3,115) = 2.901$, p = .038).

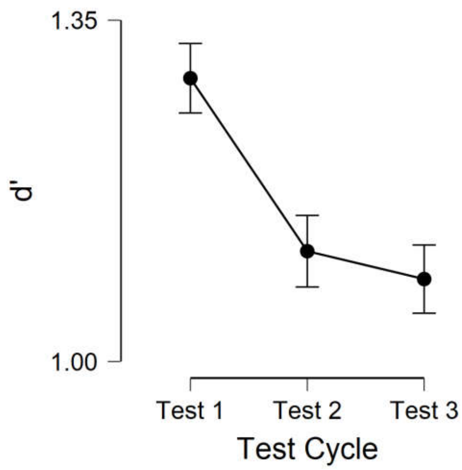


Figure 9. d' values as a function of test cycle in Experiment 1B. Error bars represent standard errors.

Table 12
A' Values in Experiment 1B

Test Cycle	Type of Inf.	Mean	SD	N
Test 1	FB+PI	0.769	0.099	30
	FBonly	0.841	0.054	30
	NoInf	0.800	0.077	30
	PIonly	0.816	0.064	29
Test 2	FB+PI	0.763	0.077	30
	FBonly	0.779	0.088	30
	NoInf	0.777	0.091	30
	PIonly	0.779	0.074	29
Test 3	FB+PI	0.752	0.089	30
	FBonly	0.787	0.075	30
	NoInf	0.762	0.077	30
	PIonly	0.781	0.090	29

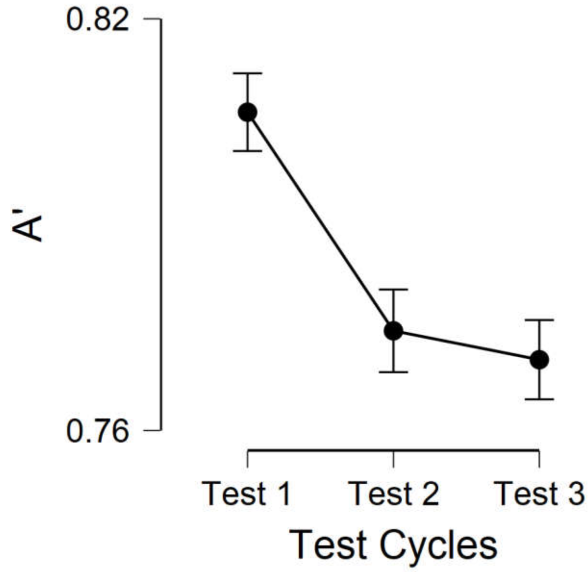


Figure 10. A' values as a function of test cycle in Experiment 1B. Error bars represent standard errors.

Response bias analysis via *C* values with 4 x 3 mixed-factor ANOVA with the same factors was conducted (Table 13). Mauchly's test of sphericity indicated a violation of sphericity assumption ($\chi^2(2) = 20.134$, $p < .001$). Therefore, following the suggestion of Field (2009, p. 461) Huynh-Feldt adjustments were made for the main effect of test cycle and interaction effects with the estimated $\varepsilon = .873$. Analysis indicated significant main effects of test cycle ($F(1.745, 200.678) = 53.02$, $MSE = 4.118$, $p < .001$, $\eta_p^2 = .260$) and type of information ($F(3, 115) = 26.61$, $MSE = 7.569$, $p < .001$, $\eta_p^2 = .410$). The interaction was also significant ($F(5.235, 200.678) = 11.95$, $MSE = 0.928$, $p < .001$, $\eta_p^2 = .176$). Pairwise comparisons of test cycles for each condition were carried out (Figure 11). In No Inf group, *C* did not differ significantly from Test 1 to Test 2 ($t(29) = 1.778$, $p = 0.240$, $d = 0.325$), from Test 1 to Test 3 ($t(29) = 0.949$, $p = 1$, $d = 0.173$), or from Test 2 to Test 3 ($t(29) = 0.649$, $p = 1$, $d = 0.119$). For PI only group, response bias was significantly more conservative in Test 1 than Test 2 ($t(28) = 4.578$, $p < .001$, $d = 0.850$) and Test 3 ($t(28)= 3.444$, $p=.002$, $d = 0.640$). However, response bias was not significantly different between Test 2 and Test 3 ($t(28) = 0.241$, $p = 1$, $d = 0.045$). For FB only group, participants were more conservative in Test 2 than Test 1 ($t(29) = 6.000$, $p < .001$, $d = 1.095$), in Test 3 than Test 1 ($t(29) = 7.203$, $p < .001$, $d = 1.315$), in Test 3 than Test 2 ($t(29) = 3.333$, $p = .004$, $d = 0.609$). In FB+PI group, likewise, C scores were higher in Test 2 than Test 1 ($t(29) = 6.730$, $p < .001$, $d = 1.229$), in Test 3 than Test 1 ($t(29) = 7.203$, $p < .001$, $d = 1.315$), and in Test 3 than Test 2 ($t(29) = 2.544$, $p = .038$, $d = 0.464$).

Table 13
C values in Experiment 1B

Test Cycle	Type of Inf	Mean	SD	N
Test 1	FB+PI	-0.016	0.295	30
	FBonly	0.095	0.315	30
	NoInf	0.259	0.333	30
	PIonly	0.100	0.263	29
Test 2	FB+PI	-0.440	0.381	30
	FBonly	-0.284	0.269	30
	NoInf	0.371	0.460	30
	PIonly	-0.193	0.344	29
Test 3	FB+PI	-0.585	0.499	30
	FBonly	-0.474	0.359	30
	NoInf	0.334	0.468	30
	PIonly	-0.179	0.409	29

ΔC values were calculated for Experiment 1B as in Experiment 1A to investigate the extent of adaptive change in response bias according to type of information (Table 14). ΔC values for FB+PI group were non-normally distributed

($D(30) = .183$, $p = .012$). The contrast analysis revealed that FB+PI group had greater shifts than PI only group ($t(86) = 2.072$, $p = .041$, $d = .494$), whereas there were not significant differences between FB+PI and FB only groups ($t(86) = 0.227$, $p = .821$, $d = .055$) or between FB only and PI only groups ($t(86) = 1.847$, $p = .068$, $d = .581$).

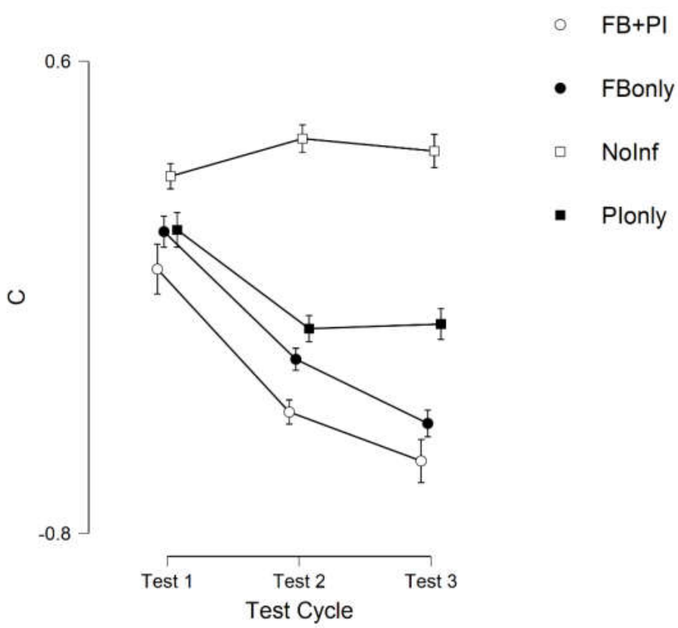


Figure 11. *C* values as a function of test cycle in Experiment 1B. FB: feedback, PI: prior information, NoInf: no information. Error bars represent standard errors.

Table 14
ΔC Values in Experiment 1B

Type of Inf.	Mean	SD	N
FB+PI	0.496	0.496	30
FBonly	0.474	0.307	30
PIonly	0.286	0.337	29

C values as a function of test block were also calculated in Experiment 1B (Figure 7).

B'' values were calculated for Experiment 1B (Table 15). The analyses, again, indicated similar patterns as with *C* values (Figure 12), except *B''* scores in Test 2 and Test 3 in FB+PI group did not differ significantly ($MD = .020$, $SE = .044$). Lastly, β values for each condition were calculated (Table 16). The optimal values for Test 2 and Test 3 were $\beta = .25$, therefore, the shifts were suboptimal as in Experiment 1A.

Table 15
B'' values in Experiment 1B

Test Cycle	Type of Inf	Mean	SD	N
Test 1	FB+PI	-0.023	0.198	30
	FBonly	0.099	0.279	30
	NoInf	0.212	0.250	30
	PIonly	0.089	0.230	29
Test 2	FB+PI	-0.249	0.213	30
	FBonly	-0.141	0.190	30
	NoInf	0.300	0.337	30
	PIonly	-0.094	0.257	29
Test 3	FB+PI	-0.268	0.234	30
	FBonly	-0.273	0.219	30
	NoInf	0.206	0.315	30
	PIonly	-0.056	0.265	29

To summarize, participants made adaptive shifts in response bias when they received supportive cues; however, they made no shifts in their criteria in the absence of supportive cues. As in Experiment 1A, PI only group did not shift their criteria from Test 2 to Test 3; however, feedback caused increasing adaptive shifts throughout the tests. Participants in the FB+PI group made greater shifts than PI only group, whereas there were no significant differences between other groups.

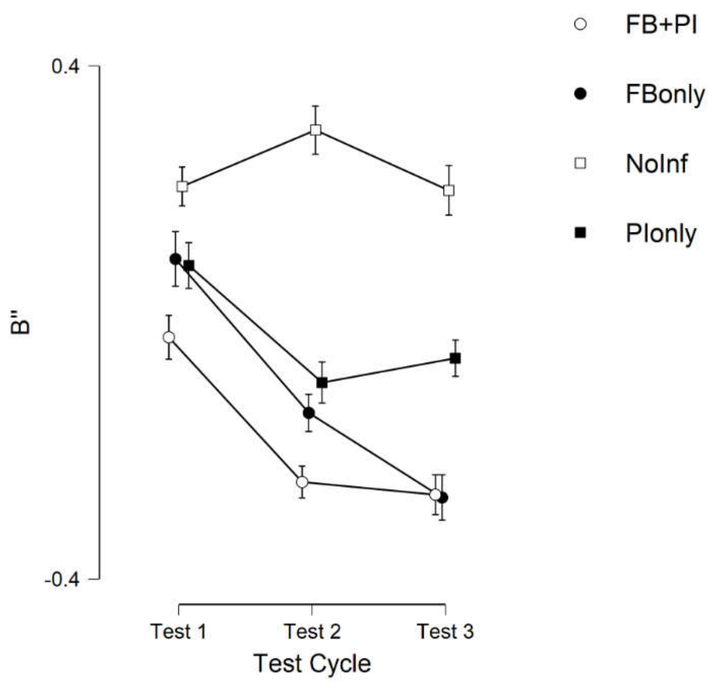


Figure 12. B'' values as a function of test cycle in Experiment 1B. FB: feedback, PI: prior information, NoInf: no information. Error bars represent standard errors.

Table 16
β *Values in Experiment 1B*

test cycle	Type of Inf.	Mean	SD	N
test 1	FB+PI	1.016	0.356	30
	FBonly	1.346	0.727	30
	NoInf	1.623	0.806	30
	PIonly	1.296	0.695	29
test 2	FB+PI	0.699	0.282	30
	FBonly	0.844	0.356	30
	NoInf	2.385	1.925	30
	PIonly	1.055	1.122	29
test 3	FB+PI	0.677	0.281	30
	FBonly	0.664	0.235	30
	NoInf	1.787	1.287	30
	PIonly	1.139	1.178	29

CHAPTER 4

DISCUSSION

In the current study, the base rates of test lists were manipulated. A shift in response bias to a more conservative position was fostered in Experiment 1A, whereas the bias was expected to become more liberal in Experiment 1B. The effects of response feedback and prior information regarding the base rates of test lists were also investigated in the experiments. The results indicated that adaptive shifts in response bias were induced with the help of supportive cues. Moreover, sensitivity decreased when the base rates of test lists were changed. In this chapter, the results of the experiments were discussed in light of the previous findings in the literature.

4.1 Sensitivity

In both experiments, sensitivity was higher in Test 1 than Test 2 and Test 3, but there were no significant differences between Test 2 and Test 3. This indicated that the decrease in sensitivity is mainly related to the manipulation of base rates of test lists. In both experiments, the first tests were standard tests in terms of base rate; therefore, they did not provide any cues in favor of any of the responses. In the absence of such cues, participants needed to rely on only their internal evidence. That situation did not require any adjustments for criterion, so *C* values for Test 1 signified their default, or initial, response bias. When base rates differed from 50%, participants had external information regarding the prior probabilities of test items. Although *d'* is assumed to be independent of criterion, a participant who integrates their internal evidence with base rate information can improve in sensitivity (see Selmeczy & Dobbins, 2013 for further discussion). In Experiment 1A and 1B, participants who were in conditions with feedback or prior information adjusted their criteria in line with the base rates of test items, yet their sensitivity decreased in Test 2 and Test 3. There are only a few experiments in which liberal or conservative conditions were compared to neutral conditions; therefore, the effect of response bias manipulation on sensitivity has not been discussed in detail in this perspective. Yet, there are findings comparable to the results of the current study. The first explanation for the decline in sensitivity is the criterion noise account (Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008). The criterion has been known to vary to some extent throughout the tasks which involve a decision process which costs memory performance. Situations that require adjustments in criterion would produce higher variability in criterion, which would decrease sensitivity more than the situations which do not require such adjustments. For instance, Kantner and Lindsay (2010, Experiment 2) carried out a base rate manipulation between 75% and 25% in their study. Participants in the feedback condition were more liberal in the 75% base rate condition compared to the control group. Moreover, their criteria were more conservative in the 25% base rate condition than the control group. More importantly, sensitivity was lower for feedback condition. Therefore, participants in the feedback group adjusted their criteria according to the base rates of test lists and making such adjustments reduced sensitivity, probably because of higher within-test variability of criterion. In another study, Benjamin et al. (2013) conducted a recognition memory task with three different response types: Yes-no recognition, 4-point scale confidence judgment, and 8-point scale confidence judgment. They found that performance became better with fewer response choices, therefore fewer criteria. This result also indicates that criterion placement is costly in terms of sensitivity. Greater number of criteria requires more adjustments, and adjustments cause worse performance. The results of this study have provided further evidence in favor of this account. However, it is also possible that participants did not take into account the supportive cues in standard lists as they did in the biased lists, because the supportive cues have no effect in criterion adjustments in these lists. This makes the recognition tasks with standard lists slightly different and easier than the tasks with biased lists. Any difference between standard and biased lists would be attributed to this difference. Yet, SDT framework does not regard the tasks with standard and biased lists as different tasks.

The second factor that would lead to a decline across test cycles is vigilance. Vigilance can also explain some variance of the decline, considering that sensitivity also decreased from Test 2 to Test 3, albeit numerically. Besides, sensitivity decreased throughout the test also in No Inf group, in which participants did not make adaptive shifts in criteria. Sensitivity does not probably stay constant throughout a test block. Macmillan and Creelman (2005, p. 46) discussed the possibility that participants sometimes have a number of trials with $d' = 0$ sensitivity. If so, these trials would disguise in the test and remain unknown in the analysis. It is reasonable to suppose that such sequences are more likely to happen towards the end of the experiments, causing lower sensitivity values for later test lists. To avoid such contamination of data, exclusion criteria can be expanded with additional measures. For instance, a case with a certain number of consecutive same responses, or a case with several responses with low response time can be excluded from the analysis. A stricter exclusion rule for sensitivity would be another choice. An alternative perspective is to keep participants concentrated via performance payoffs.

Regarding the type of information condition, sensitivity did not differ across groups in both experiments, as expected. This made it easier to interpret the effects on response bias.

4.2 Response Bias

ANOVA for *C* scores indicated that participants adapted their response bias across test cycles in accordance with the base rates of test lists when feedback or prior information were provided. By contrast, criterion remained constant in the absence of these cues. It is no surprise that base rate manipulation alone is not sufficient to induce criterion shifts (Cox & Dobbins, 2011; Koop et al., 2015; but cf. Heit et al., 2003). Participants needed additional information to benefit from prior probabilities of test trials.

In PI only condition, participants shifted their criteria in Test 2 in line with the base rates. Specifically, they became more conservative from Test 1 to Test 2 in Experiment 1A, and more liberal from Test 1 to Test 2 in Experiment 1B. Prior information's role as a reliable cue for inducing criterion shifts has been demonstrated once again in this study. An important finding is that criterion did not change between Test 2 and Test 3 in both experiments in contrast with conditions with response feedback. It shows that participants used this cue to adapt their criterion at the beginning of the test, and the degree to which prior information cue was used did not change throughout the tests. Our analysis with 3x4 test blocks also signals a slight trend in a sense that the criterion was relatively consistent across Test 2 and Test 3, especially in Experiment 1B. This is reasonable in that prior information is explicit information to be utilized in the same extent throughout the tests. On the contrary, feedback information for each response accumulates trial-by-trial to serve as a cue for base rates.

As for FB only group, criterion became significantly more conservative from Test 1 to Test 2, and from Test 2 to Test 3 in Experiment 1A. In experiment 1B, likewise, it became more liberal in every subsequent test. The experiments demonstrated that response feedback is a reliable cue for participants to make adaptive shifts in response bias in accordance with base rates. These results are consistent with some of the previous research in the literature in which criterion shifts were observed only when feedback was provided to participants (Estes & Maddox, 1995, digit and letter conditions; Kantner & Lindsay, 2010; Verde & Rotello, 2007). However, feedback was ineffective in other studies (Estes & Maddox, 1995, word condition; Hicks & Starns, 2014; Starns et al., 2010). The results were inconclusive in general regarding the studies with base rate manipulation (Estes & Maddox, 1995; Healy & Kubovy, 1977, 1978; Kantner & Lindsay, 2010; Rhodes & Jacoby, 2007). The experiments in the current study provide evidence in favor of the benefits of feedback for adaptive criterion shifts.

ΔC analyses indicated that the magnitude of criterion shifts did not differ significantly between FB only and PI only conditions, although changes were numerically greater for PI only group in Experiment 1A, and for FB only group in Experiment 1B. Once the effectiveness of these cues is established, there is no reason to assume that one of the cues is more effective than the other. However, the same analysis showed that FB+PI group had greater shifts than FB only group in Experiment 1A and PI only group in Experiment 1B. Although significant differences between groups were not found in other comparisons, all results together indicate that when feedback and prior information are used together, participants were inclined to make greater adaptive shifts.

In some studies, when criterion shifts were induced with the help of supportive cues, such as prior information, probabilistic cues, or font colors, the introduction of feedback did not change the magnitude of shifts (Healy & Kubovy, 1978; Hicks & Starns, 2014; Selmeczy & Dobbins, 2013). But the current study has shown the additional effect of feedback on prior information. Selmeczy and Dobbins (2013) considered that feedback helped participants form representations regarding the probabilities of old and new items. Therefore, it was no longer helpful when participants were already aware of the probabilities with the help of other cues. Other researchers also discussed how feedback affects response bias in terms of awareness (Koop et al., 2015; Rhodes & Jacoby, 2007). In Rhodes and Jacoby's (2007, Experiment 2) study, 25% of the participants in same-key condition were aware of the base rate manipulation, while awareness increased up to 71% in different keys condition. The magnitude of criterion shifts did not change according to awareness in same-key condition; however, shifts were greater among the participants who were aware of the manipulation in different-key condition. Importantly, participants who used different response keys made criterion shifts irrespective of their awareness (Experiment 2-3). In the current study, around a third of participants in FB only condition was aware of the manipulation, and participants shifted their criteria whether they were aware or unaware of the base rate manipulation. Considered together with Rhodes and Jacoby's findings, these results imply that feedback affects response criterion no matter participants are aware of the manipulation or not; yet, awareness also affects criterion to some extent, albeit smaller. On the other hand, some models suggest that differential reinforcement of responses results in adaptive adjustments in decision criterion (Maddox, 2002; Treisman & Williams, 1984). These models can explain the findings of Han and Dobbins (2008, 2009) in which biased feedback resulted in changes in response bias without participants being aware. In a different approach, dynamic SDT model of Turner et al. (2011) asserted a mechanism through which information on old and new distributions accumulates with the help of feedback, therefore causes adjustments in response bias. The model does not imply that participants are aware of this process and adjustments. With such processes which the models describe, feedback's effect would exceed the adjustments which are made in response to prior information. This is probably what was observed in the current study.

4.3 Conclusion and Future Directions

In the current thesis, the feasibility of criterion shifts induced by base rate manipulation was investigated. Participants made adaptive shifts in their response bias when prior information or feedback was presented; however, the base rate manipulation was not effective in the absence of these cues. The simplicity of the design helped observe the feasibility and magnitude of criterion shifts. The base rate manipulation was used in the experiments rather than strength manipulation because base rate manipulation is free from the confounds of strength manipulation method, problems of which has been discussed in Chapter 2. The base rate was manipulated between test lists rather than between blocks within a test list, or item-by-item based. Moreover, the experiment started with a standard test, then the following tests consisted of either mostly old or mostly new words. Thus, participants were expected to set their initial criterion, then made adaptive shifts only once. By this method, feedback and prior information have proven to be effective cues which foster adaptive shifts in criterion when the base rate is manipulated.

Four conditions in which supportive cues were manipulated indicated that the degree to which prior information was utilized was relatively consistent throughout tests; however, participants made greater shifts as the feedback information accumulated throughout tests. Thus, when prior information and feedback were presented together, they caused adaptive shifts in a greater degree compared with the conditions in which feedback or prior information was provided alone.

Recognition performance depends on two independent factors according to SDT: sensitivity and response bias. Although sensitivity has been investigated on many aspects in the literature, the factors which affect response bias has not been established in detail. However, as an important component of recognition memory, the decision criterion has become a major focus of research in current years (Franks & Hicks, 2016; Frithsen et al., 2017; Konkel, Selmeczy, & Dobbins, 2015; Koop et al., 2015; Koop & Criss, 2016; Selmeczy & Dobbins, 2013, 2017; Starns & Olchowski, 2015). The methods used to manipulate response bias was developed and employed in the literature. The effects of supportive cues, such as feedback, prior information, font colors and response keys, has been investigated in these studies. Yet, a lack in the literature is the absence of adequate research to explain conflicting findings. The current thesis aimed to provide further information to clarify the roles of response feedback and prior information in criterion shifts. Further research is needed to understand their roles.

In the current study, the experiments began with standard test lists; therefore, shifts from the initial criterion to liberal or conservative bias was observed. In future research, a similar method would be employed to observe within-list shifts in response bias with the help of supportive cues. It is possible to observe reliable shifts from initial (or neutral) criterion within a test list by this method. With novel methods, the limitations of adaptive shifts can be further discovered.

Individual differences in response bias have been a research topic in the following years. It was found that neutral criterion placement varies substantially among individuals (Aminoff et al., 2012; Kantner & Lindsay, 2012). Moreover, individuals vary in terms of their ability to make adaptive shifts in response bias (Selmeczy & Dobbins, 2013). It was an interesting research question whether the individuals are able to shift towards a liberal position or a conservative position from their default criterion to the same degree, or there are "liberal-inclined" and "conservative-inclined" individuals. This can be investigated by the method used in the current research.

To conclude, the current thesis has provided novel findings regarding the feasibility of criterion shifts via the base rate manipulation. Future research will allow us to understand the nature of response bias in recognition memory.

REFERENCES

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., ... Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition, 40*(7), 1016-1030. doi:10.3758/s13421-012-0204-6

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*(2), 81-99. doi:10.1037/h0029531

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language, 51*(2), 159-172. doi:10.1016/j.jml.2004.04.001

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*(1), 84-115. doi:10.1037/a0014351

Benjamin, A. S., Tullis, J. G., & Lee, J. H. (2013). Criterion noise in ratings-based recognition: Evidence from the effects of response scale length on recognition accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1601-1608. doi:10.1037/a0031849

Bruno, D., Higham, P. A., & Perfect, T. J. (2009). Global subjective memorability and the strength-based mirror effect in recognition memory. *Memory & Cognition, 37*(6), 807-818. doi:10.3758/MC.37.6.807

Cox, J. C., & Dobbins, I. G. (2011). The striking similarities between standard, distractor-free, and target-free recognition. *Memory & Cognition, 39*(6), 925-940. doi:10.3758/s13421-011-0090-3

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55*(4), 461-478. doi:10.1016/j.jml.2006.08.003

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 484-499. doi:10.1037/a0018435

Criss, A. H., Aue, W., & Kılıç, A. (2014). Age and response bias: Evidence from the strength-based mirror effect. *The Quarterly Journal of Experimental Psychology*, *67*(10), 1910-1924. doi:10.1080/17470218.2013.874037

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*(4), 316-326. doi:10.1016/j.jml.2011.02.003

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), 2-17. doi:10.1037/0278-7393.33.1.2

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1075-1095. doi:10.1037/0278-7393.21.5.1075

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Franks, B. A., & Hicks, J. L. (2016). The reliability of criterion shifting in recognition memory is task dependent. *Memory & Cognition, 44*(8), 1215-1227. doi:10.3758/s13421-016-0633-8

Frithsen, A., Kantner, J., Lopez, B. A., & Miller, M. B. (2017). Cross-task and cross-manipulation stability in shifting the decision criterion. *Memory*, *26*(5), 653-663. doi:10.1080/09658211.2017.1393090

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*(1), 8-20. doi:10.3758/BF03198438

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*(3), 546-567. doi:10.1037/0033-295X.100.3.546

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: 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

Healy, A. F., & Jones, C. (1975). Can subjects maintain a constant criterion in a memory task? *Memory & Cognition, 3*(3), 233-238. doi:10.3758/BF03212903

Healy, A. F., & Kubovy, M. (1977). A comparison of recognition memory to numerical decision: How prior probabilities affect cutoff location. *Memory & Cognition, 5*(1), 3-9. doi: 10.3758/BF03209184

Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition, 6*(5), 544-553. doi:10.3758/BF03198243

Heit, E., Brockdorff, N., & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin & Review, 10*(3), 718-723. doi:10.3758/BF03196537

Hicks, J. L., & Starns, J. J. (2014). Strength cues and blocking at test promote reliable within-list criterion shifts in recognition memory. *Memory & Cognition, 42*(5), 742-754. doi:10.3758/s13421-014-0397-y

Hilford, A., Glanzer, M., Kim, K., & Maloney, L. T. (2019). One mirror effect: The regularities of recognition memory. *Memory & Cognition*, *47*(2), 266-278. doi:10.3758/s13421-018-0864-y

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), 302-313. doi:10.1037/0278-7393.21.2.302

Hockley, W. E., & Niewiadomski, M. W. (2007). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes. *Memory & Cognition*, *35*(4), 679-688. doi:10.3758/BF03193306

Kantner, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory?. *Memory & Cognition*, *38*(4), 389-406. doi:10.3758/MC.38.4.389

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40*(8), 1163-1177. doi:10.3758/s13421-012-0226-0

Kılıç, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology, 92*, 65-86. doi:10.1016/j.cogpsych.2016.11.005

Konkel, A., Selmeczy, D., & Dobbins, I. G. (2015). They can take a hint: Older adults effectively integrate memory cues during recognition. *Psychology and aging*, *30*(4), 781-794. doi:10.1037/pag0000058

Koop, G. J., & Criss, A. H. (2016). The response dynamics of recognition memory: Sensitivity and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 671-685. doi:10.1037/xlm0000202

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. doi:10.3758/s13423-014-0703-5

Ley, R., & Long, K. (1987). A distractor-free test of recognition and false recognition. *Bulletin of the Psychonomic Society, 25*(6), 411-414. doi:10.3758/BF03334727

Ley, R., & Long, K. (1988). Distractor similarity effects in tests of discrimination recognition and distractor-free recognition. *Bulletin of the Psychonomic Society, 26*(5), 407-409. doi:10.3758/BF03334898

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. doi:10.3758/BF03212415

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

Maddox, W. T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. *Journal of The Experimental Analysis of Behavior*, *78*(3), 567-595. doi:10.1901/jeab.2002.78-567

Malmberg, K. J., Criss, A. H., Gangwani, T. H., & Shiffrin, R. M. (2012). Overcoming the negative consequences of interference from recognition memory testing. *Psychological Science*, *23*(2), 115-119. doi:10.1177/0956797611430692

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*(4), 724-760. doi:10.1037/0033-295X.105.4.734-760

Morrell, H. E., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1095-1110. doi:10.1037/0278-7393.28.6.1095

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465-494. doi:10.3758/PBR.15.3.465

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(2), 163-178. doi:10.1037/0278-7393.16.2.163

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518-535. doi:10.1037/0033-295X.99.3.518

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 305-320. doi:10.1037/0278-7393.33.2.305

Scimeca, J. M., McDonough, I. M., & Gallo, D. A. (2011). Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language, 65*(4), 363-377. doi:10.1016/j.jml.2011.04.008

Selmeczy, D., & Dobbins, I. G. (2013). Metacognitive awareness and adaptive recognition biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 678-690. doi:10.1037/a0029469

Selmeczy, D., & Dobbins, I. G. (2017). Ignoring memory hints: The stubborn influence of environmental cues on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1448-1469. doi:10.1037/xlm0000383

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145-166. doi:10.3758/BF03209391

Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition*, *34*(1), 125-137. doi:10.3758/BF03193392

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34-50. doi:10.1037/0096-3445.117.1.34

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi:10.3758/BF03207704

Starns, J. J., & Olchowski, J. E. (2015). Shifting the criterion is not the difficult part of trial-by-trial criterion shifts in recognition memory. *Memory & Cognition, 43*(1), 49-59. doi:10.3758/s13421-014-0433-y

Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language, 63*(1), 18-34. doi:10.1016/j.jml.2010.03.004

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379-1396. doi:10.1037/0278-7393.24.6.1379

Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*(5), 301-340. doi:10.1037/h0040547

Tekcan, A. I., & Göz, I. (2005). *Türkçe Kelime Normları (Turkish word norms)*. Istanbul: Bogazici Universitesi Yayinevi.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*(1), 68-111. doi:10.1037/0033-295X.91.1.68

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, *118*(4), 583-613. doi:10.1037/a0025191

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

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35*(2), 254-262. doi:10.3758/BF03193446

Wallace, W. P. (1982). Distractor-free recognition tests of memory. *The American Journal of Psychology, 95*(3), 421-440. doi:10.2307/1422134

Wallace, W. P., Sawyer, T. J., & Robertson, L. C. (1978). Distractors in recall, distractor-free recognition, and the word-frequency effect. *The American Journal of Psychology, 91*(2), 295-304. doi:10.2307/1421539

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152-176. doi:10.1037/0033-295X.114.1.152

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), 1341-1354. doi:10.1037/0278-7393.20.6.1341

APPENDICES

A: APPROVAL OF 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
www.ueam.metu.edu.tr

Sayı: 28620816 / 252

05 NİSAN 2018

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Dr. Öğretim Üyesi Aslı Kılıç ÖZHAN

Danışmanlığını yaptığınız yüksek lisans öğrencisi Cavit Deniz PALA'nın ***"Tanıma Belleğinde Temel Oran, Ön Bilgilendirme ve Geribildirimin Kriter değişimi Üzerindeki Etkisi"*** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek gerekli onay **2018-SOS-033** protokol numarası ile **16.04.2018 - 30.12.2018** tarihleri arasında geçerli olmak üzere verilmiştir.

Bilgilerinize saygılarımla sunarım.

Prof. Dr. Ş. Halil TURAN
Başkan V

Prof. Dr. Ayhan SOL
Üye

Prof. Dr. Ayhan Gürbüz DEMİR
Üye

Doç. Dr. Yaşar KONDAKÇI
Üye

Doç. Dr. Zana ÇITAK
Üye

Doç. Dr. Emre SELÇUK
Üye

Dr. Öğr. Üyesi Pınar KAYGAN
Üye

B: THE INFORMED CONSENT FORM

ARAŞTIRMAYA GÖNÜLLÜ KATILIM FORMU

Bu çalışma ODTÜ Psikoloji Bölümü Yüksek Lisans öğrencisi Cavit Deniz Pala tarafından Yard. Doç. Dr. Aslı Kılıç Özhan danışmanlığındaki yüksek lisans tezi kapsamında yürütülmektedir. Bu form sizi araştırma 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 Araştırma Laboratuvarı/Dikkat ve Bellek Laboratuvarında yapılacaktır. Üniversite öğrencileri katılımcı olarak davet edilecek, katılmak isteyenler yaklaşık 30 dakikalık bir laboratuvar seansına katılacaklardır. Çalışmada size çeşitli 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.

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 denizp@metu.edu.tr adresinden iletebilirsiniz.

Yukarıdaki bilgileri okudum ve bu çalışmaya tamamen gönüllü olarak katılıyorum.

(Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyad   Tarih   İmza

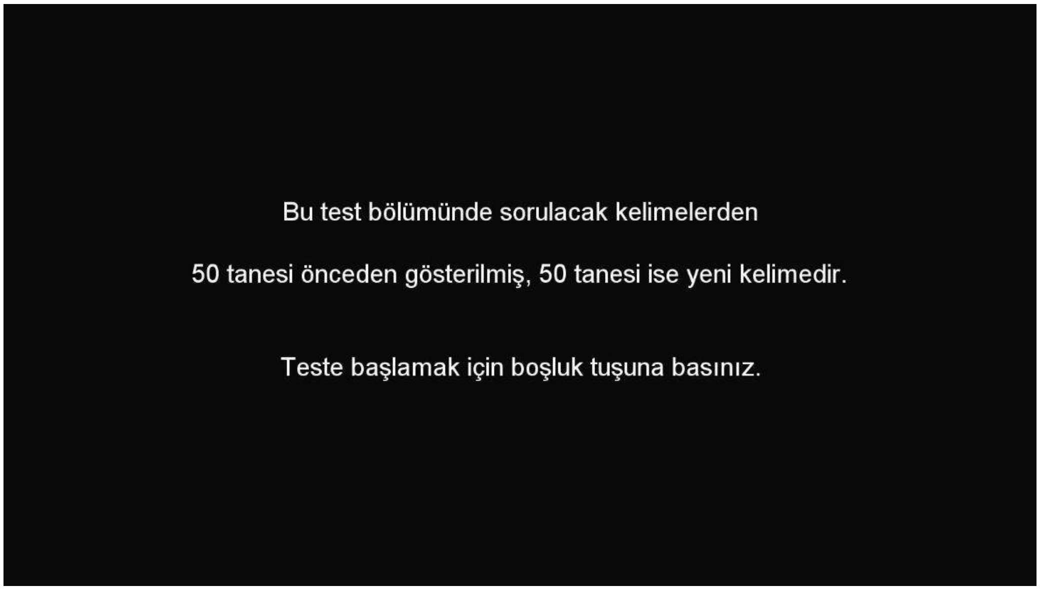
---/----/-----

C: THE INSTRUCTIONS OF THE EXPERIMENT

Merhaba, deneyimiz başlamak üzere! Lütfen yönergeyi dikkatle okuyun.
Deney 3 bölümden oluşmaktadır, her bölüm de 3 aşamadan oluşmaktadır:
çalışma, toplama, test.
Çalışma aşamasında bir dizi kelime sırayla gösterilerek aklınızda tutmaya çalışmanız istenecektir. İkinci aşamada ise ekranda sırayla gösterilecek sayıları toplamanız beklenecektedir. Ekrana gelen her bir sayıyı en son toplama ekleyerek yeni toplamı yazıp 'enter' tuşuna basacaksınız.
Örneğin ilk sayı 7 ise 7 yazarak enter tuşuna basmanız, ikinci sayı 4 ise bunu son toplam olan 7 ile toplayarak 11 sonucuna ulaşmanız beklenmektedir. Son aşamada ise ekranda sırayla kelimeler gösterilecek, her bir kelimenin o bölümün çalışma aşamasında gösterilip gösterilmediği sorulacaktır. Kelime çalışma aşamasında gösterilmiş ise 'z' tuşuna basarak 'evet', gösterilmemiş ise 'm' tuşuna basarak 'hayır' yanıtı verebilirsiniz.
Her yanıtınızdan sonra yanıtınızın doğru ya da yanlış olduğu konusunda geribildirim verilecektir. Bu aşamayla birlikte bir bölüm tamamlanacak, aynı şekilde ilerleyen 3 bölüm sonunda ise deneyimiz tamamlanacaktır.

Hazır olduğunuzda boşluk tuşuna basarak deneye başlayabilirsiniz.

D: THE PRESENTATION OF PRIOR INFORMATION



E: THE VIEW OF COMPUTER SCREEN IN TEST SESSIONS

F: THE POST-TEST QUESTIONNAIRE

1- Girdiğiniz deney üç bölümden oluşmaktaydı, her birinin ayrı bir test aşaması vardı. Bu üç test aşaması arasında herhangi bir fark dikkatinizi çekti mi?

2- Her test aşamasında bazılarını önceden çalıştığınız, bazılarını ise ilk kez gördüğünüz kelimeler soruluyordu. Test aşamalarında önceden gösterilen kelimeler ile yeni kelimelerin sayıları farklılaşıyordu, bu dikkatinizi çekti mi?

3- İlk test aşamasında gördüğünüz kelimelerin elli tanesi önceden çalıştığınız, ellisi ise yeni kelimelerdi. İkinci ve üçüncü aşamada ise yirmişer tane çalıştığınız, seksener tane ise yeni kelimeyle karşılaşıyordunuz. Bu dikkatinizi çekmiş miydi?

4- Test sırasında bir kelimeyi çalışıp çalışmadığınıza emin olmadığınızda kendinizi evet ya da hayır demeye daha yatkın hissettiniz mi?

5- Peki bu açıdan test aşamalarına göre bir farklılık oldu mu?

G: TURKISH SUMMARY / TÜRKÇE ÖZET

BÖLÜM 1

GİRİŞ

1.1 Tanıma Görevleri

Tanıma belleği daha önce karşılaşılan uyaranları yeni uyaranlardan ayırt etme becerisidir. Tipik bir evet-hayır tanıma görevi bir dizi maddenin öğrenildiği çalışma ve bu maddelerin bir kısmıyla birlikte yeni maddelerin sorulup ayırt edilmelerinin beklendiği test aşamalarından oluşur. Katılımcılar test aşamasında sorulan maddelere “evet” (“Bu maddeyi çalıştım.”) ya da “hayır” (“Bu maddeyi çalışmadım.”) yanıtı verirler. Ayrıca birçok tanıma görevinde çalışma ve test aşamalarını ayıran oyalama görevleri bulunmaktadır.

Güven kararı görevi, zorunlu seçim görevi, ilişkisel tanıma görevi, sürekli tanıma görevi, diğer başlıca tanıma belleği görevleri arasındadır.

1.2 Tanıma Belleğinde Performans Ölçümü: Sinyal Belirleme Kuramı

Bir evet-hayır tanıma görevinde katılımcıların yanıtları dört ayrı sonuç olasılığı ortaya çıkarır: Eski maddeye “evet” yanıtı “isabet”, “hayır” yanıtı “ıska” olarak adlandırılırken yeni maddeye “evet” yanıtı “yanlış alarm”, “hayır” yanıtı “doğru ret”tir. Bellek performansı isabet oranı (İO) ve yanlış alarm oranına (YAO) dayanarak ölçülür.

Sinyal Belirleme Kuramı (SBK, Green & Swets, 1966; Macmillan & Creelman, 2005) bir performans ölçüm modelidir. SBK’ye göre kişinin her test maddesine ne kadar aşina olduğunu ifade eden bir bellek gücü değeri hesaplanır. Eski ve yeni maddeler bellek gücü ölçeğinde kısmen örtüşen iki dağılım oluşturur, ancak eski dağılımı ortalama olarak daha yüksek değerdedir. Bir test maddesinin gücü katılımcının öznel olarak belirlediği bir eşiği, yani kriteri aşıyorsa o madde tanınır.

1.2.1 Duyarlılık ve tepki yanlılığı

SBK'ye göre performans iki etkene göre belirlenir: Duyarlılık, eski ve yeni maddelerin ayırt edilme düzeyi, ve tepki yanlılığı, kişinin evet yanıtı vermeye olan yatkınlığı (Macmillan & Creelman, 2005). *d'* temel duyarlılık ölçüsüdür. Sıfır değeri maddelerin ayırt edilemediğini ifade ederken duyarlılık arttıkça *d'* değeri artar. *C* ise temel tepki yanlılığı ölçüsüdür. Sıfır değeri nötr kriteri ifade ederken negatif değerler liberal (evet demeye yatkınlık), pozitif değerler ise muhafazakar (hayır demeye yatkınlık) eğilime işaret eder. Duyarlılık ve tepki yanlılığı için *A'*, *B'',* β gibi alternatif ölçüler de geliştirilmiştir (Macmillan & Creelman, 1996; Snodgrass & Corwin, 1988).

1.2.2 Tepki yanlılığı ölçümü konusundaki anlaşmazlık

Tepki yanlılığı için geliştirilen ölçüler çeşitlilik göstermektedir. Bunun ölçülerin arkasında yatan hesaplama mantığına ya da varsayımlara dayanan nedenleri olduğu gibi, tanıma belleği hakkındaki teorik bakış farklılıklarından kaynaklanan nedenleri de vardır.

Bazı deneylerde farklı koşullarda yeni dağılımlarının ortalamalarının eşit olduğu varsayılabilir, bu durumda koşullar arası YAO farkı kriter farkı olarak yorumlanabilir. Ancak YAO değerinin eşit olduğu iki koşuldan İO yüksek olan koşulun *C* değeri de yüksek bulunur. Bazı araştırmacılar bunun yanlış olduğunu öne sürmüşlerdir (Franks & Hicks, 2016; Hicks & Starns, 2014). Bu durumda YAO kriter için daha iyi bir ölçüdür. Ancak ayrıştırma modellerine (Criss, 2006, 2010; Kılıç, Criss, Malmberg, & Shiffrin, 2017; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) göre maddelerin bellek gücünün değişimlendiği çalışmalarda güçlü maddelerle birlikte test edilen yeni maddelerin bellek güçleri zayıf maddeler ile test edilenlere göre daha düşük olduğundan koşullar arası YAO farkı kriter farklılaşması anlamına gelmek zorunda değildir. Bu tür yorum farklılıkları tepki yanlılığı araştırmasında kullanılacak ölçülerin seçimi açısından kritiktir.

1.3 Kriter Değişimlemesinde Deneysel Yöntemler

1.3.1 Temel oran değişimlemesi

Karar kriterinin başlıca deneysel olarak değişimlendiği başlıca yollardan biri, test listesindeki eski ve yeni maddelerin oranlarının değiştirildiği temel oran değişimlemesidir. Test listesi çoğunlukla eski maddelerden oluştuğunda kriterin daha liberal olması beklenir, tersi çoğunlukla yeni listeler için geçerlidir. Temel oran değişimlemesi yapılan çalışmalarda çoğunlukla test listesinin temel oranının ön bilgi olarak sunulması, yanıt geribildirimi gibi yardımcı ipuçları kullanılmıştır. Yardımcı ipuçlarının kriter değişimini sağlamada gerekli olduğu gösterilmekle birlikte, her zaman yeterli olmadığı gösterilmiştir (Cox & Dobbins, 2011; Estes & Maddox, 1995; Healy & Kubovy, 1977; Koop, Criss, & Malmberg, 2015).

Healy ve Kubovy (1978) art arda gelen 8 testte temel oranları %50 ile %25 arasında değişimlemiştir. Katılımcılara temel oran ön bilgisinin sunulduğu, geribildirimin değişimlendiği deneylerde kriter temel orana uyumlu olarak değişmiştir (Healy & Jones, 1975; Healy & Kubovy, 1977 ile kıyaslayınız.). Estes ve Maddox (1995) ise 3 haneli sayılar, 3 harfli heceler ve kelimelerin uyaran olarak kullanıldığı tanıma görevlerinde temel oranı ve geribildirimi testler arası değişimlemiştir. Sayı ve hece uyaranları kullanılan görevlerde geribildirim verildiğinde temel orana göre kriter kayması gözlenmiş, ancak diğer koşullarda kriter değişmemiştir. Rhodes ve Jacoby (2007) aynı test listelerinde ekranın iki ayrı tarafında gösterilen kelimelerin temel oranlarını değişimlemiş, katılımcıların karışık sırayla sunulan test maddeleri için kriterlerini maddeden maddeye değiştirebildiklerini gözlemiştir. Bu kriter değişimi geribildirim ve yanıt tuşları ipuçlarının varlığına bağlı olarak gerçekleşmiştir. Kriter değişimini sağlayan koşullarda katılımcıların farkındalığının da arttığı, ancak değişimin farkındalıktan çok geribildirime bağlı olduğu ifade edilmiştir.

Literatürde test listelerindeki bütün maddelerin eski (yalnızca-eski) ya da yeni (yalnızca-yeni) olduğu tanıma belleği görevleri de kullanılmıştır. Bu çalışmalarda çoğunlukla yalnızca eski ya da yeni maddelerden oluşan listeler için tepki yanlılığının standart listelerden farklı olmadığı gözlenmiştir (Cox & Dobbins, 2011; Ley & Long, 1987, 1988; Wallace, 1982; Wallace, Sawyer & Robertson, 1978). Bu sonuç söz konusu listelerde bütün maddelere evet ya da hayır yanıtı vermenin testi anlamsızlaştıracağından dolayı kriterin sabit tutulmasının seçilmesinden kaynaklanmış olabilir. Koop ve arkadaşlarının (2015) geribildirim kullandıkları deneylerde yalnızca eski ve yalnızca yeni listeler ile kriter kayması sağlamaları bu yönde bir kanıttır.

1.3.2 Güç değişimlemesi

Literatürde sık kullanılan bir diğer kriter değişimleme yöntemi güç değişimlemesidir. Bu yöntemde maddelerin sunulma sıklığı (Hicks & Starns, 2014; Ratcliff, Clark, & Shiffrin, 1990, Deney 5-6; Stretch & Wixted, 1998) ya da süresinin (Bruno ve ark., 2009; Hirshman, 1995; Ratcliff ve ark., 1990, Deney 1-4) değişimlenmesi gibi yollarla maddelerin bellek güçleri farklılaştırılır. Güçlü ve zayıf maddeler farklı listelerde çalışılıp test edildiğinde güçlü listelerde daha yüksek İO ve düşük YAO gözlenir (güç temelli ayna etkisi, Glanzer & Adams, 1985; Ratcliff ve ark., 1990; ancak bakınız Hilford, Glanzer, Kim, & Maloney, 2019). Bazı araştırmacılar güç temelli ayna etkisinin kriter değişiminin sonucu olduğunu savunurken (Hicks & Starns, 2014; Starns, White, & Ratcliff, 2010; Stretch & Wixted, 1998) ayrıştırma modellerine göre bu fark eski ve yeni dağılımlarının yer değiştirmesi nedeniyledir (Criss, 2006, 2010; Kılıç ve ark., 2017; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

Güçlü ve zayıf maddelerin aynı listelerde çalışıldığı karma liste desenleri de literatürde mevcuttur. Bu çalışmaların bazılarında test maddelerinin gücüne bağlı olarak kriter değişimi sağlanamazken (Morrell Gaitan, & Wixted, 2002; Stretch & Wixted, 1998; Verde & Rotello, 2007, Deney 1-4), diğerlerinde renk ipucu (Hicks & Starns, 2014), geribildirim (Verde & Rotello, 2007) gibi yardımcı ipuçlarıyla kriter değişimi gerçekleşmiştir.

1.3.3 Alternatif yöntemler

Literatürde temel oran ve güç değişimlemelerinden başka kriter değişimleme yöntemleri de mevcuttur. Yanıt ödemeleri (response payoff, Curran, DeBuse, & Leynes, 2007; Healy & Kubovy, 1978; Van Zandt, 2000), olasılıksal ipuçları (Selmeczy & Dobbins, 2013, 2017), yanlı geribildirim (Han & Dobbins, 2008, 2009) bu yöntemler arasındadır.

Han ve Dobbins (2008, 2009) yanlış alarmlara "doğru" geribildiriminin verildiği "liberal-yanlı geribildirim" ile ıskalara "doğru" geribildiriminin verildiği "muhafazakar-eğilimli geribildirim" koşullarını değişimlemiş, katılımcıların kriterlerini koşulla uyumlu olarak değiştirdiklerini raporlamıştır. Katılımcılar kriterlerini geribildirim koşuluna göre değiştirmelerine rağmen geribildirimdeki yanlılığın farkında olmamışlardır.

Selmeczy ve Dobbins (2013) ise test aşamasında bir sonraki test maddesinin eski olma olasılığının sunulduğu ipuçları yoluyla kriter değişimi sağlamayı başarmışlardır. Katılımcılar "muhtemelen eski" ipuçlu maddeler ile "muhtemelen yeni" ipuçlu maddeler arasında maddeden maddeye kriter kayması gerçekleştirmişlerdir. Bu deneylerde kriter kaymasının büyüklüğü de ipucunun güvenilirliğine bağlı olarak değişmiştir.

1.4 Yardımcı İpuçlarının Tepki Yanlılığı Üzerindeki Etkisi

Özetlendiği gibi, tepki yanlılığının değişimlendiği çalışmalarda geribildirim ve ön bilgi yardımcı ipuçları olarak kullanılmıştır. Bu çalışmalarda geribildirim ve ön bilginin etkililiği deney desenlerine göre farklılık göstermiştir. Ön bilgi kullanımının kriter değişimi sağlamada yeterli olmadığı çalışmalar olduğu gibi (Cox & Dobbins, 2011; Healy & Jones, 1975; Healy & Kubovy, 1977; Wallace, 1982) ön bilgi ile kriter değişimi gerçekleştirilen çalışmalar da mevcuttur (Aminoff et al., 2012; Franks & Hicks, 2016; Frithsen, Kantner, Lopez, & Miller, 2017; Healy & Kubovy, 1978). Aynı şekilde, geribildirimin kriter değişimi sağlamada tek başına yeterli olduğu çalışmalar mevcut olduğu gibi (Estes & Maddox, 1995; Koop, et al., 2015; Rhodes & Jacoby, 2007, Verde & Rotello, 2007) yeterli olmadığı çalışmalar da vardır (Hicks & Starns, 2014; Starns ve ark., 2010).

1.5 Tepki Yanlılığında Çoklu Değişimlerin Sağlanabilirliği

Literatürdeki birçok çalışma liberal ve muhafazakar koşulların art arda değişimiyle kriterde çoklu değişimleri sağlamayı amaçlamıştır. Bu çalışmaların bazılarında koşullar arasında anlamlı bir kriter farkı gözlenememişken (Healy & Jones, 1975, Healy & Kubovy, 1977) diğerlerinde kriter kayması sağlanmıştır (Aminoff ve ark., 2012; Franks & Hicks, 2016; Healy & Kubovy, 1978). Ayrıca, test blokları uzun olduğunda karar kriterinin ilk test bloğunun koşulundan etkilendiği gözlenmiştir (Hicks & Starns, 2014; Verde & Rotello, 2007). İlk test bloğunun etkisinin sonraki koşullara taşındığı Han ve Dobbins'in (2008, 2009) deneylerinde de gösterilmiştir. Bunun yanında, söz konusu deneylerde ilk blok liberal ya da muhafazakar bir koşula sahip olmuş, böylece katılımcıların başlangıç kriterlerinin ne olduğu, ilk koşula uyum sağlamak için ne düzeyde bir değişim gerçekleştirdikleri tespit edilememiştir. Bu tezde standart test listesiyle başlayıp liberal ya da çoğunlukla eski ya da çoğunlukla yeni test listeleriyle devam eden deneyler gerçekleştirilmiş, böylece katılımcıların başlangıç kriterleri ve adaptif kriter kaymasının miktarı ölçülebilmiştir.

1.6 Mevcut Çalışma

Mevcut tezin ilk amacı test listelerinin değişimlenme sıralarından kaynaklanan karıştırıcı etkilerin kontrol edildiği bir deneysel desen ortaya koymaktır. İkinci amacı ise temel oran değişimlemesi yapılan bir tanıma görevinde geribildirim ve ön bilginin kriter kaymasının miktarı (ya da gerçekleşmesi) üzerindeki etkisini incelemektir.

Mevcut çalışmada katılımcılar Türkçe kelimeler ile evet-hayır tanıma görevleri gerçekleştirmişlerdir. Deney 1A'da üç çalışma-test döngüsü varken test listelerinin temel oranları sırayla %50, %20 ve %20'dir. Böylece katılımcıların nötrden liberal kritere kaymalarının düzeyleri incelenmiştir. Deney 1B'de ise temel oranlar sırayla %50, %80 ve %80 olarak değişimlenmiş, böylece nötrden muhafazakar kritere geçiş gözlenmiştir. Deneylerde ayrıca ön bilgi ve geribildirimin varlığı 4 grupta değişimlenerek etkileri incelenmiştir. Ek olarak, Deney 1A'da deney sonrası anketi ile katılımcıların temel oranların farkında olup olmadığı tespit edilmiştir. Duyarlılık ve tepki yanlılığı ölçüleri olarak *d'* ve *C* kullanılmıştır.

Bu çalışma temel oran değişimlemesinin kullanıldığı bir desende ön bilgi ve geribildirimin ortogonal olarak değişimlendiği ilk çalışmadır. Bu yöntem özellikle geribildirimin etki mekanizmasını anlamak için önemlidir. Bazı araştırmacılar geribildirimin etkisini farkındalık üzerinden tartışmışken (Hicks & Starns, 2014; Koop ve ark., 2015; Rhodes & Jacoby, 2007; Selmeczy & Dobbins, 2013) diğerleri geribildirimin koşullar arasında ayırıcı pekiştirme gerçekleştirmesi yoluyla etkide bulunduğunu öne sürmüşlerdir (Han & Dobbins, 2008, 2009; Maddox, 2002; Treisman & Williams, 1984). Ön bilginin sunulduğu koşullarda geribildirimin etkisinin olup olmaması ve geribildirimin sunulduğu koşullarda katılımcıların temel oranların farkında olup olmaması, alternatif yorumları sınama fırsatı vermiştir.

Ön bilgi ve geribildirimin yokluğunda kriter değişiminin olmaması, bu yardımcı ipuçlarının sunulduğu koşulda ise standart listelerden yanlı listelere geçişte kriterin değişmesi öngörülmüştür. Duyarlılığın ise testler arasında farklılık göstermemesi beklenmiştir.

BÖLÜM 2

YÖNTEM

2.1 Katılımcılar

Orta Doğu Teknik Üniversitesinden 120 öğrenci Deney 1A, ayrıca 124 öğrenci Deney 1B'ye ekstra ders notu karşılığında katılmıştır. Kötü performans gösteren (görev boyunca *d'* < .5, ya da bir test döngüsünde *d'* < .1) ya da tepki yanlılığı uç değerlerde olan (*C* değeri ilgili test döngüsünün ortalamasının 3 standart sapma değerinden daha yüksek ya da daha düşük) kişiler analize dahil edilmemiş, böylece analiz Deney 1A için 115, Deney 1B 119 katılımcı ile gerçekleşmiştir.

2.2 Uyaranlar

Deney 1A 300 çalışma ve 210 yeni test kelimeleriyle, Deney 1B 300 çalışma ve 90 yeni test kelimeleriyle gerçekleştirilmiştir. Kelimeler Türkçe Kelime Normları'ndan (Tekcan & Göz, 2005) rastgele seçilmiştir.

2.3 Desen

Deney 1A 3 çalışma-test döngüsünden oluşmuştur. Katılımcılar her bir döngüde 100 kelime çalışmış ve 100 kelime ile test edilmişlerdir. Testlerin temel oranları sırayla %50, %20 ve %20'dir. Ayrıca bilgi tipi katılımcılar arası 4 koşulda değişimlenmiştir: Bilgi yok (BYok), Yalnızca ön bilgi (ÖB) yalnızca geribildirim (GB) ve geribildirim ile ön bilgi (GB+ÖB). Ön bilgi alan gruplarda test aşamalarından önce o testin temel oran bilgisi verilmiş, temel oran test boyunca ekranda sunulmuştur. Geribildirim alan gruplarda her yanıtın ardından geribildirim verilmiştir. Deney 1A'da aynı zamanda katılımcıların temel oranların farkında olup olmadıklarını ölçen bir test sonrası anketi uygulanmıştır.

Deney 1B iki fark dışında her bakımdan Deney 1A ile aynıdır: Test listelerinin temel oranları sırayla %50, %80 ve %80'dir, ve test sonrası anketi uygulanmamıştır.

2.4 Prosedür

Deneyler MATLAB üzerinde geliştirilen bir yazılım ile gerçekleştirilmiştir. Katılımcılar koşullara bilgisayar tarafından atanmıştır. Bilgilendirilmiş onam formu imzalandıktan ve yönerge sunulduktan sonra deney başlamıştır. Çalışma aşamalarında 100 kelime sırayla koyu gri arkaplan üzerine beyaz renkli Arial 65 font ile sunulmuştur. Kelimeler 1 saniye süreyle ve 100 milisaniye arayla sunulmuştur. Çalışma aşamasını 45 saniyelik toplama görevi takip etmiş, sonrasında test aşaması başlamıştır. Test kelimeleri sunulurken kelimelerin daha önce sunulup sunulmadığı sorulmuş, ön bilgi ekranın sağ üst köşesinde sarı font ile sunulmuştur. Yanıtlar klavye ile verilmiş ve yazılım tarafından kaydedilmiştir. Deney 1A'daki katılımcılar deney sonrası anketini sözel olarak yanıtlamışlardır.

BÖLÜM 3

SONUÇLAR

3.1 Veri Transformasyonu ve Atılması

Her koşul için İO ve YAO hesaplanmış ve gerekli düzeltmeler yapılmıştır. Daha sonra *d'* ve *C* hesaplanarak 4 x 3 karma ANOVA ile R, JASP ve SPSS yazılımlarıyla analiz edilmiştir. Temel oranlar arası kriter kayması düzeyini belirten Δ*C* değerleri hesaplanmış ve ÖB, GB ve GB+ÖB grupları arasında planlı contrast analizleri ile kıyaslanmıştır. Analizlerin normallik, varyansların homojenliği ve küresellik varsayımları kontrol edilmiş, post-hoc kıyaslamalarda *p* değerleri için Bonferroni düzeltmesi yapılmıştır.

3.2 Deney 1A

3.2.1 Duyarlılık

Duyarlılık, *d'* değerleri ile yapılan 4 x 3 karma ANOVA ile analiz edilmiştir. Bilgi tipi (BYok, ÖB, GB, GB+ÖB) denekler-arası, test döngüsü (1, 2, 3) denek-içi değişkendir. Yapılan analizde test döngüsü temel etkisi anlamlı ($F(2, 222) = 8.559$, $MSE = 1.255$, $p < .001$, $\eta_p^2 = .072$), bilgi tipi temel etkisi anlamsız ($F(3, 111) = 0.387$, $MSE = 0.244$, $p = .762$, $\eta_p^2 = .010$) bulunmuştur. Koşullar arası anlamlı bir etkileşim yoktur ($F(6, 222) = 0.599$, $MSE = 0.088$, $p =.731$, $\eta_p^2 = .016$). İkili kıyaslamalara göre duyarlılık Test 1'den ($M = 1.408$, $SD = 0.481$) Test 2'ye ($M = 1.265$, $SD = 0.568$, $t(114) = 2.898$, $p = .014$, $d = 0.270$) ve Test 3'e ($M = 1.204$, SD

= 0.597, $t(114) = 4.303$, $p < .001$, $d = 0.401$) düşmüştür. Ancak, Test 2 ve Test 3 arasında anlamlı bir duyarlılık farkı bulunmamıştır ($t(114) = 1.129$, $p = .784$, $d = 0.105$). *A'* ile yapılan duyarlılık analizi aynı örüntüyü ortaya koymuştur.

3.2.2 Tepki yanlılığı

Tepki yanlılığı *C* ile 4 x 3 karma ANOVA ile analiz edilmiştir.

Analize göre test döngüsü temel etkisi ($F(2, 222) = 202.352$, $MSE = 5.433$, $p <.001$, $\eta_p^2 = .480$), bilgi tipi temel etkisi ($F(3, 111) = 7.345$, $MSE = 1.626$, $p < .001$, $\eta_p^2 = .166$), ve etkileşim etkisi $F(6, 222) = 9.391$, MSE = 0.498, $p <.001$, $\eta_p^2 = .202$) anlamlıdır. İkili kıyaslamalara göre BYok grubunda Test 1 ile Test 2 (29)= 0.180, $p = 1$, $d = 0.033$), Test 1 ile Test 3 ($t(29)= 1.397$, $p = .498$, $d = 0.255$), ya da Test 2 ile Test 3 ($t(29)= 1.426$, $p = .477$, $d = 0.260$) arasında tepki yanlılığı farkı yoktur. ÖB grubunda kriter Test 1'e kıyasla Test 2'de ($t(27)= 5.968$, $p < .001$, $d = 1.128$) ve Test 3'te ($t(27)= 7.738$, $p < .001$, $d = 1.462$) daha muhafazakardır. Ancak kriter Test 2 ile Test 3 arasında farklılık göstermemiştir ($t(27)= 2.268$, $p = .076$, $d = 0.429$). GB grubunda ise kriter Test 1'den Test 2'ye ($t(26)= 4.422$, $p < .001$, $d = 0.851$), Test 1'den Test 3'e ($t(26)= 7.091$, $p < .001$, $d = 1.365$), ve Test 2'den Test 3'e ($t(26)= 3.246$, $p = .005$, $d = 0.625$) daha muhafazakar hale gelmiştir. GB+ÖB grubunda da kriter aynı şekilde Test 1'den Test 2'ye ($t(29)= 7.934$, $p < .001$, $d = 1.449$), Test 1'den Test 3'e ($t(29)= 10.270$, $p < .001$, $d = 1.875$), ve Test 2'den Test 3'e ($t(29)= 3.019$, $p = .001$, $d = 0.551$) farklılık göstermiştir. *B''* analizleri C ile aynı sonuçları ortaya koymuştur.

Kriter kaymasının boyutunun bilgi tipinden etkilenip etkilenmediğini incelemek için ÖB, GB ve GB+ÖB grupları için ΔC [($C_{Test3} + C_{Test2}$) / 2 – C_{Test1}] hesaplanmıştır. Kontrast analizlerine göre GB+ÖB grubu GB grubuna göre daha fazla değişim göstermiş ($t(82)=2.276$, $p=.025$, $d = 0.625$), ancak değişim düzeyi açısından ÖB grubundan farklılaşmamıştır ($t(82)=1.524$, $p=.131$, $d = 0.412$). ÖB ile GB grupları arasında da kriter kayması açısından anlamlı bir fark bulunmamıştır ($t(82)=0.754$, $p=.453$, $d = 0.192$).

3.2.3 Test Sonrası Anketi

Katılımcıların test listelerinin temel oranları hakkındaki farkındalığı test sonrası anketi ile ölçülmüştür. Test 2 ve Test 3'te maddelerin çoğunun eski olduğunu, ancak Test 1'de olmadığını fark eden katılımcılar farkında olarak sınıflanmıştır.

BYok grubundaki 30 katılımcıdan 2'si, ÖB grubundaki 28 katılımcıdan 25'i, ve GB+ÖB grubundaki 30 katılımcıdan 28'inin temel oranların farkında olduğu tespit edilmiştir. GB grubundaki katılımcıların 8'i temel oranları doğru şekilde tahmin edebilmişken 17'si edememiş, 2 katılımcı ise net bir yanıt verememiştir. ΔC analizine göre farkında olan (M = 0.476, SD = 0.354) ve olmayan (M = 0.353, SD = 0.337) katılımcılar kriter değişimi gerçekleştirmiş, ancak değişim düzeyi açısından gruplar arasında anlamlı bir fark gözlenmemiştir ($t(23)$ = 0.836, p = . 412, d = .358). Buna göre, yalnızca geribildirim, katılımcıların temel oranların farkında olup olmamasından bağımsız olarak kriter kaymasını sağlamada yeterli olmuştur.

3.3 Deney 1B

3.3.1 Duyarlılık

d' analizi test döngüsünün duyarlılık üzerinde anlamlı etkisini ortaya koymuştur ($F(2, 230) = 9.808$, MSE = 1.481, $p < .001$, $\eta_p^2 = .079$), ancak bilgi tipinin anlamlı bir etkisi bulunmamıştır ($F(3, 115) = 1.863$, MSE = 0.920, p = .140, η_p^2 = .046). Koşullar arası etkileşim de gözlenmemiştir ($F(6, 230) = 1.321$, MSE = 0.200, p = .248, η_p^2 = .033). Deney 1A'daki gibi, duyarlılığın Test 1'den (M = 1.290, SD = 0.519) Test 2'ye (M = 1.113, SD = 0.534, $t(118) = 3.431$, p = .002, d = .315) ve Test 3'e (M = 1.085, SD = 0.504) düştüğü ($t(118) = 4.181$, $p < .001$, d = .383); ancak Test 2 ile Test 3 arasında farklılık göstermediği tespit edilmiştir ($t(118) = 0.563$, p = 1, d = .052). *A'* analizleri de *d'* sonuçlarına paraleldir.

3.3.2 Tepki yanlılığı

C analizine göre test döngüsü ($F(1.745, 200.678) = 53.02$, MSE = 4.118, p < .001, η_p^2 = .260) ve bilgi tipinin ($F(3, 115) = 26.61$, MSE = 7.569, $p < .001$, η_p^2 = .410) tepki yanlılığı üzerinde anlamlı etkileri bulunmuştur; etkileşim etkisi de aynı şekilde anlamlıdır ($F(5.235, 200.678) = 11.95$, $MSE = 0.928$, $p < .001$, $\eta_p^2 = .176$). BYok grubunda kriter Test 1 ile Test 2 ($t(29) = 1.778$, $p = 0.240$, $d = 0.325$), Test 1 ile Test 3 ($t(29) = 0.949$, $p = 1$, $d = 0.173$), ya da Test 2 ile Test 3 ($t(29) = 0.649$, $p = 1$, $d = 0.119$) arasında anlamlı olarak değişmemiştir. ÖB grubunda kriter Test 1'de Test 2'ye ($t(28) = 4.578$, $p < .001$, $d = 0.850$) ve Test 3'e ($t(28)= 3.444$, $p=.002$, $d = 0.640$) göre daha muhafazakardır. Ancak, Test 2 ile Test 3 arasında anlamlı bir kriter farkı yoktur ($t(28) = 0.241$, $p = 1$, $d = 0.045$). GB grubunda ise, Katılımcılar Test 1'den Test 2'ye ($t(29) = 6.000$, $p < .001$, $d = 1.095$), ve Test 3'e ($t(29) = 7.203$, $p < .001$, $d = 1.315$), aynı şekilde Test 2'den Test 3'e ($t(29) = 3.333$, $p = .004$, $d = 0.609$) adha liberal olmuşlardır. GB+ÖB grubunda da aynı şekilde kriter Test 1'den Test 2'ye ($t(29) = 6.730$, $p < .001$, $d = 1.229$), Test 1'den Test 3'e ($t(29) = 7.203$, $p < .001$, $d = 1.315$), ve Test 2'den Test 3'e ($t(29) = 2.544$, $p = .038$, $d = 0.464$) giderek düşmüştür. *B''* analizlerinin sonuçları da *C* ile aynı örüntüyü sergilemiştir.

ΔC ile yapılan kontrast analizine göre GB+ÖB grubu ÖB grubundan daha çok kriter kayması gerçekleştirmiştir ($t(86) = 2.072$, $p = .041$, $d = .494$). Bununla birlikte, GB+ÖB ile GB ($t(86) = 0.227$, $p= .821$, $d = .055$) ya da GB ile ÖB grupları ($t(86) = 1.847$, $p = .068$, $d = .581$) arasında adaptif değişim düzeyi açısından bir fark görülmemiştir.

BÖLÜM 4

TARTIŞMA

4.1 Duyarlılık

İki deneyde de duyarlılık Test 1'den Test 2 ve 3'e düşüş göstermiş, ancak Test 2 ve 3 arasında anlamlı bir fark gözlenmemiştir. Bu durum duyarlılıktaki düşüşün temel orandaki değişimlemeye bağlı olduğuna işaret etmektedir.

Kriter gürültüsü açıklaması (Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008) bu bulguyu açıklayabilir. Bu açıklamaya göre kriter, verilen her kararda belirli bir miktar değişmektedir. Kriterde ayarlamalar yapmayı gerektiren durumlarda bu değişimin miktarı artarak duyarlılıkta düşüşe neden olabilir (Benjamin, Tullis, & Lee, 2013; Kantner & Lindsay, 2010). Standart testler kişinin kriterinde bir ayarlama yapmasını gerektirmezken çoğu eski ya da çoğu yeni maddelerden oluşan listeler bu ayarlamayı gerektirmiş, bu da duyarlılığın düşmesine neden olmuş olabilir. Katılımcıların deney boyunca uyanıklıklarında yaşanan düşüş de duyarlılıktaki düşüşün bir kısmını açıklamaktadır. Duyarlılık gruplar arasında bir farklılık göstermemiş, bu sayede tepki yanlılığındaki etkileri yorumlamak kolaylaşmıştır.

4.2 Tepki Yanlılığı

Deneylerde temel oran değişimlemesi, geribildirim ve ön bilginin sunulduğu koşullarda tepki yanlılığında farklılaşmaya yol açmış, ancak yardımcı ipuçlarının yokluğunda kriter değişimi gerçekleşmemiştir.

ÖB koşulunda kriter Test 1'den Test 2 ve 3'e farklık göstermiş, ancak Test 2 ile Test 3 arasında anlamlı bir farklılık ortaya çıkmamıştır. Bu bulgu katılımcıların ön bilgiden yararlanma miktarlarında 2. ve 3. Testler boyunca tutarlılık sergilediklerine işaret etmektedir.

Deney 1A'da geribildirim verilen koşullarda kriter Test 1, 2 ve 3 boyunca sürekli daha muhafazakar hale gelmiştir. Deney 1B'de ise benzer bir şekilde katılımcılar testler boyunca gittikçe daha liberal hale gelmişlerdir. Bu çalışma geribildirimin kriter değişimini sağlamada tek başına yeterli olduğu yönünde kanıt sağlamaktadır. ΔC analizine göre ise geribildirim ve ön bilginin bir arada sunulduğu koşullardaki kriter adaptasyonu miktarı yalnızca geribildirim ya da yalnızca ön bilgi sunulan koşullara göre daha yüksek olma eğilimi göstermiştir. Bazı araştırmacılar tarafından geribildirimin temel oran hakkındaki farkındalığı artırması yoluyla etkisi tartışılmıştır (Koop et al., 2015; Rhodes & Jacoby, 2007; Selmeczy & Dobbins, 2013). Bu çalışmada geribildirimin temel oranların farkında olmayan katılımcılarda da kriter değişimine yol açtığı, aynı zamanda ön bilginin sunulduğu koşulda da geribildirimin kriter değişim miktarını artırdığı gözlenmiştir. Bu bulgular geribildirimin etkisinin farkındalıktan başka etki yolları olduğunu ortaya koymaktadır. Bazı modeller geribildirimin yanıtların ayırıcı pekiştirmesi yoluyla tepki yanlılığını değiştirdiğini öne sürmüştür (Maddox, 2002; Treisman & Williams, 1984). Bu modeler mevcut bulguları ve Han ve Dobbins'in (2008, 2009) bulgularını açıklayabilmektedir. Turner, Van Zandt ve Brown (2011) geribildirimin eski ve yeni dağılımları hakkında bilgi sağlayarak tepki yanlılığına etki ettiğini öneren bir model ortaya koymuşlardır. Bu model de bulguları açıklaması mümkündür.

4.3 Sonuç ve Gelecek Çalışmalar

Mevcut tezde temel oran değişimlemesiyle kriter kaymasının sağlanabilirliği incelenmiştir. Temel oran değişimlemesinin ancak geribildirim ve ön bilgi yardımıyla sağlanabildiği gözlenmiştir. Uygulanan deneysel desenin basitliği sayesinde kriterde büyük miktarda kaymalar gözlenebilmiştir. Çalışma, ön bilgi ve geribildirimin etki yolları hakkında veriler ortaya koymuştur.

Deneyler standart testlerle başlamış, arkasından yanlı test listeleri uygulanmıştır. Böylece katılımcıların başlangıç kriterleri ve kriterlerindeki adaptif kayma miktarları ölçülebilmiştir. Gelecek çalışmalarda benzer bir desenle aynı test içinde bloklar arası kriter kayması miktarı ölçülebilir. Benzer bir deneysel desenle kişilerin nötrden liberale ve nötrden muhafazakara adaptir değişim gerçekleştirebilme düzeyleri arasındaki ilişkiler incelenebilir, kişiler arası farklılıklar araştırılabilir. Gelecek çalışmalar tanıma belleğinde tepki yanlılığı konusunun anlaşılmasına katkı sağlayacaktır.

H: TEZ İZİN FORMU / THESIS PERMISSION FORM

<u>**ENSTİTÜ /** INSTITUTE</u>

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 □

<u>**YAZARIN /** AUTHOR</u>

Soyadı / Surname : Pala

Adı / Name : Cavit Deniz

Bölümü / Department : Psikoloji / Psychology

<u>**TEZİN ADI /** TITLE OF THE THESIS</u> (**İngilizce** / English) : The Roles of Feedback and Prior Information on The Effectiveness of The Base Rate Manipulation for Response Bias

<u>**TEZİN TÜRÜ**</u> **/** <u>DEGREE**:**</u> **Yüksek Lisans** / Master ■ **Doktora** / PhD □

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ■

2. **Tez <u>iki yıl</u> süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **<u>two years</u>. *** □

3. **Tez <u>altı ay</u> süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **<u>six months</u>**. ***** □

***** *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 ………………