8-1-2012

Output Interference and Strength Based Mirror Effect in Recognition Memory

Asli Kilic
Syracuse University, akilic@syr.edu

Abstract

The strength based mirror effect (SBME) refers to higher hit rates and lower false alarm rates for strongly encoded items. The SBME has been explained by two alternative mechanisms: *differentiation* and *criterion shift*. The differentiation account posits that as the memory traces are strengthened, the memory for the items is less noisy and therefore are less confusable with the stored memory traces. The criterion shift account, on the other hand, suggests that the tendency to endorse a test item differs between strong and weak test lists. When participants receive a test after studying a strong list of items, they require more evidence to endorse an item. This account suggests that the SBME is observed due to the decision processes rather than memory. One strong piece of evidence for the criterion shift account is that the SBME is observed after studying a mixed list in which half of the items are strengthened and the other half are not. The differentiation mechanism can not account for the SBME observed after a mixed study list. However, recent findings of output interference (OI) in recognition memory have been explained by encoding at retrieval, which suggest that differentiation can also take place at retrieval. This suggests a third possible explanation for the SBME might be encoding during retrieval. In a series of experiments, we investigated the effect of strengthening items at study on OI in pure and mixed list paradigms. The results from the pure list paradigm (Experiment 1) showed both OI and the SBME as well as an interaction between OI and the SBME, as the item-noise models predicted. The results from the mixed list paradigm (Experiment 2-4) showed that the SBME was observed only when participants were informed of the strength of the list that they would be tested on. This finding adds constraints to the criterion shift account which states that participants shift their criterion to endorse a test item according to the type of the test list and that shift in the criterion causes the SBME. These results suggest that both criterion shift and differentiation account could be playing a role in the SBME after studying pure lists, and that only criterion shift account plays a role in SBME after studying a mixed list, but only when

participants are explicitly aware of the testing conditions.

OUTPUT INTERFERENCE AND STRENGTH BASED MIRROR EFFECT IN RECOGNITION MEMORY

by

Aslı Kılıç

B.S., Bilkent University, 2004
M.S., Middle East Technical University, 2007

Dissertation
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philospohy
in Experimental Psychology.

Syracuse University
August 2012

Contents

List of Tables

List of Figures

Introduction

The term episodic memory was first introduced by Tulving (1972), referring to a type of memory that stores temporal-spatial information (Tulving, 1983). Tulving further argued that episodic memory is susceptible to interference as the act of retrieval from episodic memory can change the stored information by adding new information to memory. In a broad sense, the aim of this project was to investigate the mechanisms of episodic memory by focusing on how the information stored in memory can interfere with testing and how testing can interfere with the information already stored in memory.

In the laboratory, episodic memory can be tested by a recognition task. In a recognition task, participants are given a list of items to study, and later they are given a test list that contains the studied items and new ones. Their task is to endorse the studied items (targets) and reject the new items (foils). From the signal detection theory (SDT), memory evidence for targets and foils is assumed to be normally distributed with a greater mean evidence for targets (Green & Swets, 1966). At test, participants set a criterion and if the memory evidence of the test item exceeds that criterion, then the item is endorsed if it does not, it is rejected (Figure 1). Thus, stemming from SDT, two distinct processes contribute to item recognition judgments. One is the memory process which is measured by the memory evidence accumulated for the test item; the other being a meta-cognitive process which modulates the placement of the criterion.

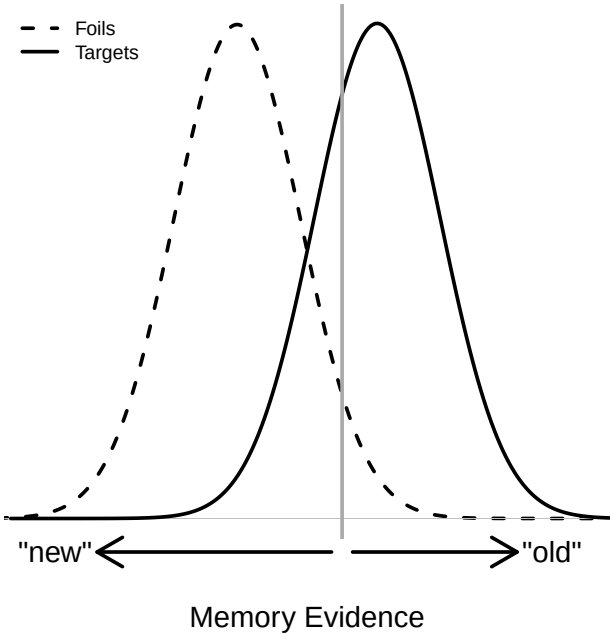

Figure 1: Illustration of the signal detection theory. Grey line represents the criterion set by the participant. If memory evidence exceeds criterion test item is judged to be 'old' otherwise 'new'.

Sources of Interference In Recognition Memory

There is an ongoing debate on the relative contribution of item and context information to the memory evidence in item recognition. Item information is described as the semantic, perceptual or orthographic properties of the items whereas the context is described as the temporal-spatial information (Estes, 1955; Howard & Kahana, 2002) or information regarding the environment (Malmberg & Shiffrin, 2005; Murnane, Phelps, & Malmberg, 1999). *Context-noise models* assume that the memory evidence is based on the similarity between the test context and the previous contexts in which the test item has been encountered (Bind Cue Decide Model of Episodic Memory, BCDMEM; Dennis & Humphreys, 2001). Thus, in BCDMEM, the noise in recognition memory is the previously encountered contexts of the test item. In the model, when a word is encoded, the current

study context is bound to the studied word. At the beginning of the test, the study context is reinstated and for each test item, a context vector is retrieved which is a set of context features. This set of context features includes all of the pre-experimental contexts in which the word was encountered, as well as the latest study context for the targets. The recognition decision is made based on a match between the reinstated context and the retrieved context. When the value of this match exceeds a criterion, the item is endorsed. Since the retrieved context of the foils does not include the study context, the value of the match between the reinstated context and the retrieved context of a foil is less likely to exceed the criterion compared to that of a target. It is clear from the mechanisms of BCDMEM that in context-noise models, context information is the only factor that contributes to memory evidence while the memory traces of other items have no influence on the recognition decision. Therefore, according to BCDMEM, interference in recognition memory can only be caused by the past contexts which are retrieved at the time of the test.

Alternatively, *item-noise models* propose that both context and item information contribute to memory evidence (D. L. Hintzman, 1984; D. Hintzman & Ludlam, 1980; McClelland & Chappell, 1998; Murdock, 1982, 1983, 1993; Shiffrin & Steyvers, 1997). For example, in Retrieving from Memory Model (REM; Shiffrin & Steyvers, 1997), items are represented as vectors of feature values with both item and context information. During study, when a memory trace is stored, both item and context features of that trace are stored probabilistically. During retrieval, the test item features are matched to every trace in a given context and when the global match exceeds a criterion, the item is endorsed. For example, the targets are more likely to show a greater global match because if the test item is a target, it is more likely to match one of the memory traces. When memory is probed with a foil, it matches the traces in memory less well. The foil matches traces stored in memory due to item features that are shared with other memory traces. Thus, the critical factor for a recognition decision in this model is that the memory evidence is influenced by the match between the test item and the traces of the other items in memory. According to

these models, context also plays an important role in item recognition, such that the memory search set is defined according to the match between the context features of the test item and memory. However, the emphasis is on the influence of other traces in the memory search set.

To evaluate the item-noise models and the context-noise models, we will use the strength-based mirror effect and the output interference effect observed in recognition memory. Previous research on the item-noise models shows a unified explanation for these two effects. In this project, we will further evaluate the predictions from the item-noise models by using REM to provide a theoretical explanation and we will use the diffusion model (DM; Ratcliff, 1978) to test the predictions from the alternative accounts. In the following sections, REM and DM, will be described in detail. Then, the strength based mirror effect and the output interference effect will be reviewed along with the predictions from REM and the other alternative accounts. Finally, these predictions will be tested in four experiments.

Retrieving Effectively from Memory

In REM, both item and context information are represented as vectors of feature values which are positive integers. In the current simulations, we will focus on only the item information and assume that the context information is the same for all of the items presented in the same study list (i.e., REM.4 Shiffrin & Steyvers, 1997). Thus, the context vector will not be added to the memory trace and as a result will not be used for retrieval (Criss, 2006; Criss, Malmberg, & Shiffrin, 2011). By doing so, only the item-noise mechanism of REM will be tested throughout this project. The length of the vector for each item is set to 20 and each feature value is sampled independent of one another from the geometric distribution as follows:

$$P(v) = (1-g)^{v-1}g, \quad v = 1, ..., \infty, \tag{1}$$

where g is the parameter for the geometric distribution and v is the actual sampled value for each element in the vector. For the item features, g has been set to 0.35.

One of the basic assumptions in REM is that the items are stored in memory probabilistically, which means that the memory traces of the items are not complete and they are error-prone. For the study, each feature is stored with some probability (u). If a feature is not stored, that feature value in the trace is zero which indicates a lack of information. For each feature value stored in memory, another parameter moderates the probability of correctly copying the feature value ($c = 0.70$). If a feature is not correctly copied, then a random value which is sampled from the same geometric distribution is then stored in the trace. This satisfies the error-prone assumption of encoding. For each additional encoding of a given item, only the features that were zero are replaced with a probability of u and the features that are already stored do not change. The parameter u varies because it is reflects encoding accuracy which is determined by the experimental design and the participant. Therefore in the literature, u is typically fit separately to each experimental condition or fit to data participant group. In the three experiments here, we fit u to the strong and weak conditions.

In REM, at retrieval, the test item is matched to all of the traces in memory and a subjective likelihood is calculated from the following equation:

$$\lambda_{(i,j)} = (1-c)^{nq_{(i,j)}} \prod_{v=1}^{\infty} \left[\frac{c + (1-c)g(1-g)^{v-1}}{g(1|-g)^{v-1}} \right]^{nm_(v,i,j)}, \tag{2}$$

where j indexes the test item, i indexes the memory trace, c is the probability of correctly copying the feature value, nq is the number of non-zero feature mismatches, v is the feature value sampled from the geometric distribution and nm is the number of non-zero feature matches. As shown by the equation, the feature values of zero do not contribute to the

subjective likelihood as they are uninformative.

To make a decision, the subjective likelihood ratios are averaged across traces stored in memory. The odds (Φ) is calculated as follows:

$$\Phi_j = \frac{1}{n} \sum_{i=1}^{n} \lambda_{(i,j)}, \tag{3}$$

where n is the number of traces in the memory search set, i indexes the memory trace and j indexes the test item. If Φ exceeds the *criterion*, item j is judged to be 'old', if Φ is below the *criterion*, then the item j is judged to be 'new'. From SDT perspective, Φ can be considered as the memory evidence and *criterion* is the threshold for endorsing an item (see Figure 1). In REM, the feature selection, storage, and comparison process is repeated for multiple simulated participants, producing variability in the data. Just as with human data, the average of the simulated participants is plotted in the figures to represent the model fit.

Although REM is a process model that produces the target and the foil distributions along a memory evidence scale ($\log(\Phi)$) and can account for accuracy data observed in recognition experiments, the version that will be used in this project does not have a mechanism to account for reaction time data [1]. In order to test the predictions regarding the reaction time data, we will use the diffusion model, which will be discussed in detail in the next section.

The Diffusion Model

The DM can be considered a dynamic detection model which takes accuracy, reaction time and the reaction time distributions into account and uses these in an attempt to untangle the underlying psychological processes (e.g., memory and meta-cognitive decision processes; Ratcliff, 1978). In the DM, each choice is represented as a threshold for a

[1] See Diller, Nobel, and Shiffrin (2001); Malmberg (2008) for a version of REM that predicts reaction time data

decision to be made. For example, in a 'yes/no' recognition task, the two response boundaries (thresholds) are 'old (yes)' and 'new (no)' (Figure 2). Once the test item is presented, the memory evidence regarding the item will accumulate over time towards one of the response boundaries. This rate of accumulation, driven by the quality of evidence, is determined by the *drift rate* parameter (v) in the DM. At each time step, the memory evidence is sampled and compared to a criterion which is also known as the *drift criterion* (Ratcliff, 1978, 1985; Ratcliff & McKoon, 2008; Ratcliff, Van Zandt, & McKoon, 1999). If the sampled evidence exceeds that criterion, the evidence accumulates towards the 'old' boundary, otherwise the evidence accumulates towards the 'new' boundary. As a result of this continuous sampling process, memory evidence reaches one of the response boundaries, which determines the decision.

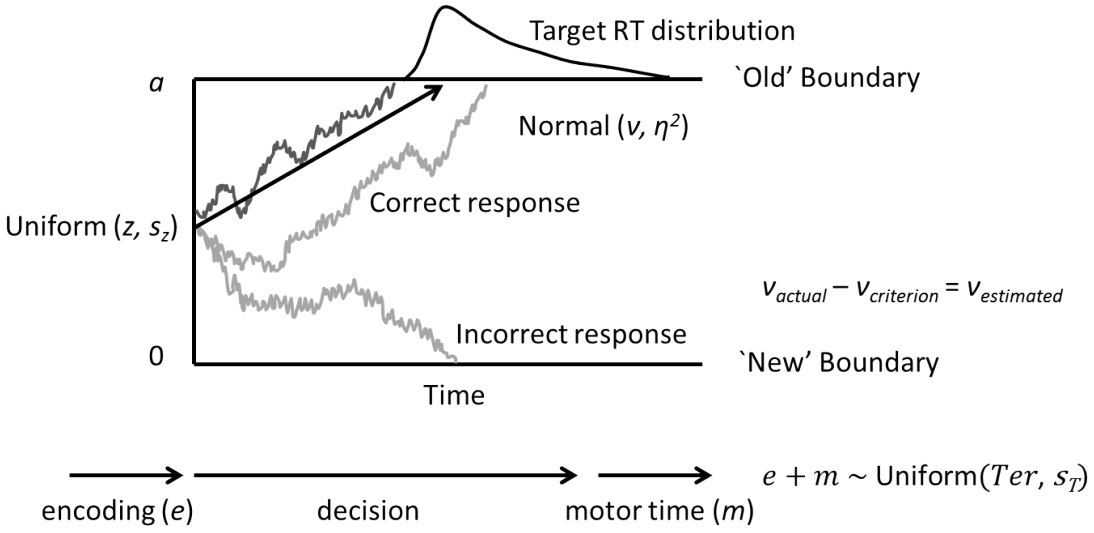

Figure 2: Illustration of the diffusion model shows the accumulation paths for correct and incorrect responses. In recognition memory, the upper boundary represents 'old' responses and the lower boundary represents 'new' responses. In this specific example, the accumulation of evidence for targets is illustrated.

In a recognition task, the memory evidence has been defined as the match between the test item and memory. It should be mentioned that the DM does not provide an explanation of how the match between the test item and memory is calculated at the same

level of detail as REM does. This is because REM is a process model that describes the mechanisms underlying memory whereas the DM is a descriptive model that measures the contribution of cognitive components in any two-choice decision. When the test item is a target, the evidence sampled at each time step is likely to have a greater match value compared to the drift criterion. Thus, the target items tend to reach the 'old' boundary more often. Similarly, since the foil items do not match well with memory, the match value tends to fall below the criterion and evidence is more likely to accumulate towards the 'new' boundary. The target items tend to produce positive drift rates and the foil items tend to produce negative drift rates because the drift criterion is typically set to the zero point drift rate. Thus, only the relative position of the drift criterion can be defined and the drift criterion cannot be specified independent of the mean drift rates. Additionally, the absolute value of the drift rate determines the quality of information such that a greater absolute value of v represents faster and more accurate responses and a lower absolute value of v represents slower and less accurate responses. The drift rate is assumed to be normally distributed within trials with a mean of ξ and a standard deviation of s. s is a scaling parameter and usually fixed to the arbitrary value of 0.1 (Ratcliff, 1978; Ratcliff et al., 1999; Vandekerckhove & Tuerlinckx, 2007). Furthermore, ξ is assumed to be normally distributed with a mean of v and a standard deviation of η. Therefore, v is the mean drift rate across trials and η is the across trial standard deviation for the drift rate.

Another parameter which is called the *boundary separation* (a), is well known to characterize the speed-accuracy trade-off (Ratcliff, 1985; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). When the response boundaries are narrow, meaning that a is small, the evidence reaches the boundaries faster, which is observed as a shorter reaction time. Since less evidence is required to reach a boundary and the evidence is noisy, there are more instances of the evidence accumulating towards the incorrect boundary. On the other hand, when the boundaries are wider, more evidence is required to accumulate in order to reach a boundary and response time is slower overall. For example, if the initial accumulation of an

item is towards the incorrect boundary, it is possible that the accumulation path can change its direction over time and drift towards the correct boundary. One should note that the change in the parameter a is under the control of the participant. For example, if participants are instructed to give correct responses without any time limitation during the test, they would set wider boundaries. On the other hand, if they are required to give fast responses and allowed to sacrifice accuracy, they will need to set the boundaries narrower and as a result, require less evidence to judge whether an item is old or new (Ratcliff & McKoon, 2008). Therefore, the boundary separation parameter is described as reflecting the participant's cautiousness.

The second parameter which also moderates a meta-cognitive process, is the *starting point* (z) that takes a value between 0 and a. z represents the point between the two boundaries at which the accumulation of evidence starts. This parameter is in accordance with the criterion in SDT in a way that the changes in z accounts for the same effects that are accounted for by the criterion in SDT (Criss, 2010). For example, Criss (2010) showed that z was the only parameter that could explain the changes in the response bias when participants were given different proportions of targets in a test list. The starting point is also assumed to be distributed uniformly with a range of s_z. Finally, the non-decision component which refers to the time required to encode the item and execute a motor response is modelled in a uniform distribution with a mean of T_{er} and a range of s_T.

The Strength Based Mirror Effect

In a strength based mirror effect (SBME) paradigm, the items are strengthened during the study either by increasing the study time, repetitions or manipulating the depth of encoding. When strength is manipulated across lists, for example, participants study a list of items which are all strengthened, the probability to endorse a target (hit) increases and the probability to endorse a foil (false alarm) decreases (Cary & Reder, 2003; Criss, 2006,

2009, 2010; Glanzer & Adams, 1985; Starns, White, & Ratcliff, 2010; Stretch & Wixted, 1998) for the strong list. As the memory evidence of the strong targets are greater, the hit rates for the strong targets are higher than the hit rates for the weak targets. The critical finding in this paradigm is that the false alarm rate for the strong foils (foils that are tested along with the strong targets) is lower than the false alarm rate for the weak foils (foils that are tested along with the weak targets). Hence, the debate on the mechanisms that explain the SBME stems from why encoding strength would affect the unstudied items during test.

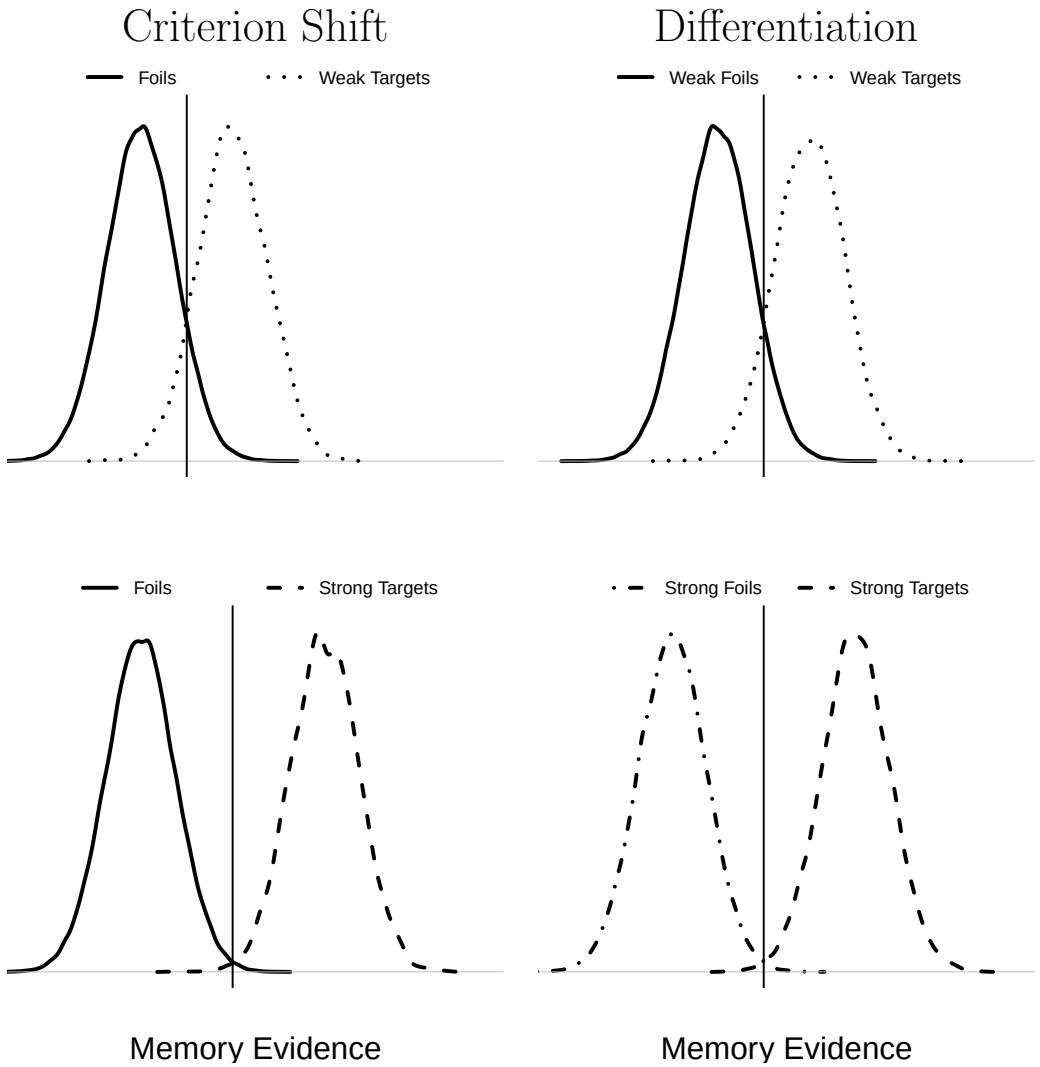

Figure 3: Illustration of the two theoretical explanations for the strength based mirror effect.

One of the theoretical explanation of the SBME is known as the *differentiation* account. The term differentiation has been first defined by E. J. Gibson (1940) in a verbal

learning task as a decrease in the associations across paired-stimuli when the associations within the pairs are strengthened. Later, the idea of differentiation has been applied to the perceptual identification tasks (J. J. Gibson & Gibson, 1955). J. J. Gibson and Gibson (1955) suggested that perceptual learning is the ability to discriminate between stimuli as they become more distinct from each other.

The mechanism of differentiation has been first introduced in memory research by Shiffrin, Ratcliff, and Clark (1990) to account for the null list-strength effect observed in recognition memory (Ratcliff, Clark, & Shiffrin, 1990). The list-strength effect, a predecessor of the SBME, refers to decrease in accuracy for the weak items that were studied along with strengthened items in a free recall task (Tulvig & Hastie, 1972). When Ratcliff et al. (1990) strengthened all the items in a list (pure condition) and strengthened only a subset of items in a list (mixed condition) to investigate the effect of strengthening on the accuracy of weak items, they did not observe a difference in the recognition performance on the weak items. However, the global matching models (e.g., Gillund & Shiffrin, 1984; Murdock, 1982) predicted a decrease in accuracy for the weak items that are studied in a mixed-strength list as these models propose that the recognition decision is based on the match between the test item and all the items stored in memory. Thus, the match between a weak item and other stronger items in memory (mixed condition) will be noisier than the match between a weak item and other weak items in memory (pure condition). This mechanism will cause a decrease in accuracy for the weak items that are presented in a mixed list. However, the results did not show a difference in the accuracy for the weak items across pure and mixed list conditions in recognition. Shiffrin et al. (1990) implemented the differentiation mechanism in the search of associative in memory (SAM) model by assuming that strengthening increase the associations between the item and its memory trace as well as between the item and context. Then, the increase in these associations would decrease the relative association between the item and other items' memory traces, and this assumption causes the stronger items to differentiate from the

other items in memory.

The differentiation mechanism from SAM was brought to the REM model (Shiffrin & Steyvers, 1997) and later used to explain the SBME (Criss, 2006) by proposing that when items are strengthened during study, the memory of these items interferes less with the foils. As a result, the strong foils are less confusable with the contents of episodic memory in comparison to the weak foils. This mechanism is inherent in REM, as items are strengthened during study, the memory traces of those items are updated by replacing the zero features. Hence, the memory traces become more complete and informative after strengthening. During retrieval, as the test item is matched to every trace, strong targets match better with their traces as they have fewer non-zero features compared to the traces for weak targets. This increase in the match produces a higher value of subjective likelihood ratio and as a result, a higher value of the odds (Φ). As a result, the mean memory evidence of a strong target distribution is greater than the mean memory evidence of a weak target distribution (see Figure 3 for an illustration). The foils on the other hand, match poorly with the memory traces of the strong targets because the traces of those items are more complete and dissimilar to the test item. The dissimilarity between the features of the test item and the non-zero features of the traces will produce lower values of subjective likelihood ratios. Thus, the lower values of subjective likelihood ratios are averaged over traces to produce the odds which end up with a lower value. As a result, REM predicts a lower mean memory evidence for strong foils compared to the mean memory evidence for weak foils (Criss, 2006, 2009, 2010). A similar differentiation mechanism is also shared with the Subjective Likelihood Model (SLiM; Criss & McClelland, 2006; McClelland & Chappell, 1998).

An alternative account for the SBME is based on the meta-cognitive processes that suggests a shift in the criterion placement (Starns, Ratcliff, & White, 2012; Stretch & Wixted, 1998; Verde & Rotello, 2007). According to this account, similar to REM, the memory evidence distribution of the strong targets shifts due to the increased memory

evidence obtained from strengthening the targets during study. However, this account suggests that the memory evidence of the foils which are presented along with either the strong or the weak targets are distributed similarly. Thus, the mean memory evidence is similar for the strong and the weak foils as these items are not studied in the experiment and there is no reason to have different memory evidence distributions for the foils. However, the difference in the false alarm rates across foils that is found in empirically documented needs to be accounted for, and this account proposes that participants set a more stringent criterion for the strong lists (Figure 3). As a result, the false alarms will be lower for the strong foils because of the change in the criterion and the hit rates will be higher for the strong targets as the target distribution shifts towards greater values of memory evidence. The criterion shift account is derived from the assumption that when participants know that they are being tested on strong items, they require more memory evidence to endorse a test item because they expect to have stronger memories for targets (Cary & Reder, 2003; Hirshman, 1995; Starns et al., 2012; Verde & Rotello, 2007). This explanation for the SBME has also been implemented in BCDMEM by matching the learning parameter at study and test (Starns et al., 2010). The learning parameter modulates the probability of binding the current study context to targets. Thus, to simulate strengthening at study, the learning parameter is increased and an increase in the learning parameter represents a stronger link between the target and the current study context. In BCDMEM, an estimate of the learning parameter is used at retrieval to calculate the likelihood ratio of the probability of observing the data given that the item is a target to the probability of observing the data given that the item is a foil. An increase in the estimated learning parameter decreases the likelihood of endorsing a foil item as the expected amount of information known about the test item increases. As a result, the model predicts lower false alarm rates for the strong foils. It is important to note that the estimated learning parameter at test is not related to encoding during the test, which has not been implemented in BCDMEM, but it is an estimate of how well the information is

encoded during study. Thus, this mechanism is not differentiation but adapting a different criterion for the strong test lists.

A final possible account of the SBME can be due to an encoding-at-retrieval mechanism. Recently, the encoding-at-retrieval mechanism has been implemented in REM to explain output interference in recognition memory (Criss et al., 2011). In the following sections both output interference and the explanations for output interference will be reviewed in detail along with a unified explanation of the SBME and output interference in REM. First, we will review the DM applications for the predictions of the item-noise models and the criterion shift account of the SBME.

The Diffusion Model Analysis in the SBME

The SBME has been extended to reaction time data by applying the DM in two recent studies (Criss, 2010; Starns et al., 2012). Criss (2010) applied the DM to test the two possible explanations of the SBME: the differentiation account and the criterion shift account. The differentiation account posits that the mean of the memory evidence distribution is lower for the strong foils compared to that of the weak foils. This shift in the foil distribution is assumed to be due to the decreased match between the foils and the more complete traces in memory. For example, REM predicts low odds value (Φ) for strong foils compared to the weak foils. Thus, this explanation predicts a faster and more accurate responses for the strong targets and the strong foils. Then, the absolute value of the drift rates are expected to be greater for the strong items compared to the absolute value of the drift rates for the weak items, both targets and foils.

Instead of a shift in the foil distribution, the criterion shift account suggests that the criterion is more stringent for the strong lists and more evidence is required to endorse an item in a strong list. This account can be implemented in the DM by assuming that participants shift their starting point towards the 'new' response boundary for strong lists. In order to investigate these predictions, Criss (2010) manipulated word frequency and list

strength orthogonally in a single item recognition experiment (Experiment 2). The response accuracy and the reaction time data was fit to three models: the drift rate model, the starting point model, and the mixed model. In the drift rate model, the drift rate parameter was allowed to be different for the strong targets, weak targets, strong foils, and weak foils. In the starting point model, the drift rates were fixed for the foil conditions as a function of the list strength, but the starting points and the target drift rates were allowed to vary across two list strength conditions. The final model was a mixed model which combined the first two models. In this model, the starting point parameter was allowed to change across list strength in addition to a change in the drift rate across targets and foils separately as a function of list strength. Criss (2010) reported that the best fitting model to the SBME paradigm was the mixed model, in which the drift rate was lower in absolute value for the weak foils than for the strong foils in support of predictions made by the differentiation account. The starting point for the strong list was found to be closer to the 'old' boundary which showed that participants were faster and more likely to respond 'old' in the strong test list. This finding is contrary to the criterion shift account which posits that participants shift their criterion to be more stringent in endorsing a strong target.

Another possible explanation of the criterion shift account in the DM is the shift in the drift criterion. The drift criterion shift account states that when participants know that they will be tested only on strong items, they would require a better match between the test item and memory to accumulate the evidence towards the 'old' response boundary. Then, the changes in the drift rates of the foils can be reinterpreted as a change in the drift criterion. For example, the shift in the drift criterion will result in a difference in the drift rate for strong foils because the strong foils are less likely to exceed the criterion in comparison to the weak foils and drift towards the 'new' boundary faster. Even though this interpretation is based on the decision processes, it could still account for the differences in the drift rates for the foil distributions. One should also note that the exact placement of the drift criterion cannot be measured because the drift criterion is the zero

point drift rate in the DM and the values of the drift rate parameter of each condition is relative to the drift criterion of that condition. In that sense, drift criterion resembles the criterion in SDT (Ratcliff, 1985; Starns et al., 2012).

In a recent study, Starns et al. (2012) investigated the drift criterion shift and the differentiation account of the SBME. List strength was manipulated in two different study conditions. In the pure-study condition, participants studied either strong words (repeated five times) or weak words (presented only once) in separate lists and a test followed each list. In the mixed-study condition, in each study list, half of the words were strong and the other half were weak. In this condition, the subsequent test list was either composed of only the strong targets or only the weak targets along with foils. The mixed study list condition is critical for testing the differentiation account because this account does not predict the SBME in a mixed-study condition. Memory should not differ between foils tested on a strong and a weak test because the match between the foils and memory traces encoded during a mixed-study list is the same for both strong and weak lists. Thus, if the SBME is observed in a mixed-study list paradigm, then it is due to the criterion shift and decision processes rather than a change in the memory evidence. Results from the accuracy data showed the SBME in the mixed-study list condition as well as in the pure-study list condition.

(Starns et al., 2012) fit the DM, in which the drift rates and the starting points were allowed to vary across test list strength, to the data. From the fits, the difference in the starting point parameter estimates suggested that the participants started from a point closer to the 'old' response boundary, which was a close replication of what Criss (2010) showed. The starting point parameter had an opposite pattern of what the criterion shift account predicts. The model fits from the pure-study list condition also replicated the findings from Criss (2010) study. The most critical finding from the DM fit was the difference in the drift rate parameter values observed in the mixed-study list condition. The absolute value of the drift rates were higher for strong targets and foils compared to

absolute value of the drift rates of weak targets and foils. This finding was also consistent with the recognition accuracy data that showed the SBME. Starns et al. (2012) suggested that this difference in the drift rates was due the changes in the drift criterion. As the global match between a foil and a mixed strength study list would be similar across test lists, the SBME observed after a mixed-study list cannot be explained by differentiation. Thus, one can question what causes a change in the drift rate parameter values for foils after studying lists with pure strength. Is the increase in the absolute value of the drift rate caused by a more stringent drift criterion or differentiation? Unfortunately, the source of the change in the drift rates of the foils cannot be separated in the DM.

There are two important points about this mixed-study list paradigm that will be addressed in the current project. The first one is that in the Starns et al. (2012) study, participants were told about the strength of the targets being tested. Thus, they were let to adopt a different criterion for different strength conditions of the test list. In the current project, the effect of informing participants of the strength of the tested items will be evaluated in a mixed-study list paradigm. The SBME is expected to be observed only when the participants are informed as they are essentially told to change their criterion.

The second point is to evaluate whether the SBME is predicted due to differentiation at retrieval. We will test this hypothesis in this project but first we will review the output interference effect and how REM explains this effect by implementing an encoding at retrieval mechanism. We will discuss whether this added mechanism can produce differentiation at retrieval and as a result the SBME in a mixed-study list paradigm.

Output Interference

The output interference (OI) effect refers to a decrease in memory accuracy towards the end of the test list. OI has been found in paired-associate paradigms (Roediger & Schmidt, 1980; Tulving & Arbuckle, 1963, 1966; Wickens, Born, & Allen, 1963), category cued recall

paradigms (Roediger & Schmidt, 1980) and also in recognition tasks (Criss et al., 2011; Malmberg, Criss, Gangwani, & Shiffrin, 2012; Murdock & Anderson, 1975; Ratcliff & Murdock, 1976). Criss et al. (2011) demonstrated that in a 'yes/no' recognition task, the decrease in accuracy was mainly caused by the decrease in hit rates as a function of test position. Additionally, OI has been observed in alternative forced choice (AFC) recognition tasks which eliminate the response bias. In these AFC tasks, participants are given a target item with either one foil (2AFC) or three foils (4AFC) and they are asked to pick the target. Thus, in these tasks, the decision process such as the tendency towards one of the responses as in 'yes/no' recognition, does not play a role on the responses given by the participant because they must chose a target on each trial. OI observed in AFC tasks suggest that decrease in the accuracy is not related to a possible change in the response bias across the test.

One of the three possible explanations for OI proposes that the source of the interference is the memory trace of the other items at retrieval. REM can account for OI if encoding is included during test (Criss et al., 2011). Encoding during test has been implemented in other memory models (e.g. SAM, Gillund & Shiffrin, 1984; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980, 1981) before. However, REM is the first model that uses encoding at retrieval to explain OI in recognition memory. The assumption for encoding at test is as follows: If a test item is judged to be old, the best matching trace in memory is updated. If a test item is judged to be new, a new memory trace is stored. This assumption causes errors in the memory storage. Similar to updating at study, when any trace is updated at test, only the missing features are updated by copying the features of the test item probabilistically. For example, a foil item can match well to the memory traces especially when these traces are not complete and this match elicits an old response (false alarm). When a foil item is called 'old' the best matching trace will be incorrectly updated with the feature values of the foil. Then, the probability of a match between the trace and the subsequent test items (e.g., even the target item from

which the memory trace was created) will decrease. Thus, all the incorrectly endorsed foil items update an incorrect trace which decreases the probability of a correct match between a target and its trace. Even updating a trace for a correctly endorsed target will decrease the hit rate for subsequently tested items via differentiation (e.g., the updated target trace will match all future test items less well). Likewise, the foils or the targets that are judged to be 'new' are stored in a new memory trace which, in return, increases the set size resulting in an increase in the interference for subsequent test items. Both updating and adding a new trace decrease the hit rate for the subsequent test items.

Opposite effects are predicted in the false alarm rate as a result of updating and adding new traces. When the memory traces are updated, the traces will store more features and as a result, the probability of a match between a subsequent foil and a trace decreases. On the other hand, when a test item is stored as a new incomplete trace, the probability of a match between a foil and a trace increases because of the increased noise in memory. In other words, updating memory traces results in a differentiation in which the false alarm rate decreases, adding a new memory trace results in an increase in the false alarm rate. In a typical recognition task that has been evaluated so far, because of the balance between updating and storing a new trace, the hit rate always goes down and the false alarm rate tends to stay the same across test positions.

A second possible explanation for OI comes from the context-noise models in which only the context information contributes to the recognition decision (Dennis & Humphreys, 2001). Criss et al. (2011) speculated on how BCDMEM can account for OI by suggesting that the decrease in the memory evidence as a function of test position could be due to the change in the reinstated context over the course of the test. This hypothesis is based on the assumption that context drifts as a function of time or input (Estes, 1955; Howard & Kahana, 2002). If the reinstated context drifts during the test, then the match between the reinstated context and the retrieved context declines. However, this explanation has been challenged by a couple findings presented in Criss et al. (2011) Experiment 2. The authors

showed that OI is robust to a 20-minute delay between study and test sessions and the decrease in accuracy over the 20-minute delay is much lower than the decrease in accuracy over the course of the test. If the reinstated context changes as a function of time, then the decrease in accuracy over the 20-minute delay should have been greater than the decrease in the accuracy over about 3-minutes (based on the test duration). Additionally, OI was still observed when the study-test lag was held constant, meaning that items were tested in the same order as they were presented. Thus, the retention interval was the same across all of the test trials and OI was observed when the study-test lag did not confound with the test position. Thus, in order to explain OI by a change in the reinstated context, the drift in the context should be attributed to the input received rather than time over the course of the test. A recent study of a probed recall task showed that probing memory with an item can change the context at test and facilitate recall of neighboring study items as they share similar study context (Kiliç, Criss, & Howard, 2012). Thus, one can assume that retrieving an item can cause a change in the reinstated context as the retrieved information of the item influences the reinstated context. As a result, the reinstated context will be less similar to the study context due to the interference caused by previously tested items. This will result in a poor match between the reinstated context and the retrieved context of the current test item, and thus will decrease both hit rates and the false alarm rates towards the end of the test list. However, this mechanism is not inherent in BCDMEM. Finally, the real challenge comes from a recent finding which shows a release from OI as a result of the category change during test (Malmberg et al., 2012). Malmberg et al. (2012) presented subjects with sets of words from two different categories in random order. At test, all participants received a 2AFC test with pairs composed of targets and foils with the same category. For half of the participants the order of the testing was random and for the other half, the test list was blocked with respect to the categories. The results showed a release from OI in the blocked condition such that the accuracy increased when the category was changed in the middle of the test. These results suggest that a change in the nature of the

test items causes a release from the build up. These findings are challenging for the context-noise models because the similarity between the memory traces should not affect the recognition decision according to BCDMEM.

Yet another possible explanation for OI could be a speed-accuracy trade-off due to a decrease in attention towards the end of the test session. It is also possible that being tested on successive trials causes some sort of an habituation to test trials. In the proactive interference literature, a topic of debate has been whether the processes at encoding or retrieval cause the decrease in performance. For example, Watkins and Watkins (1975) defined the contributing processes at encoding as habituation of successive trials and further tested whether a change in the nature of the stimuli would cause a "perceptual alerting" and as a result benefit encoding. They defined these changes in the processes at encoding as an explanation for the release observed in proactive interference when the nature of the items has been changed. The alternative account of the release from proactive interference was the decrease in interference at retrieval. In this project, we extended the definition of the habituation process to retrieval and further defined decrease in attention as habituation, fatigue or boredom over the course of the test session. More generally, the attention hypothesis refers to a decrease in recognition performance that is explained by any processes other than the interference from the memory traces of recently tested items. As a result, the attention hypothesis suggests that the accuracy goes down due to faster and less accurate responses rather than a decrease in the memory evidence. This hypothesis can be tested by looking at the reaction times in addition to memory accuracy as a function of test position. Since neither the version of REM used in this project nor BCDMEM accounts for reaction time data, the DM will be used to evaluate whether the decrease in accuracy is due to the speed-accuracy trade-off or due to the decrease in memory evidence across test positions.

The Diffusion Model Predictions for OI

Item noise models and the attention hypothesis have distinct predictions in the DM regarding OI. Item-noise models explain OI due to the encoding at retrieval mechanism which increases the interference as the test progresses, and thus predicts a decrease in memory evidence towards the end of the test list. The attention hypothesis, on the other hand, proposes that accuracy decreases due to strategic processes, and predicts lower accuracy and slower reaction times towards the end of the test list.

As the drift rate parameter moderates the rate of memory evidence accumulation, item-noise models predict a decrease in drift rates as a function of the test position. A decrease in the drift rate is expected especially for the targets as Criss et al. (2011) showed a reduction in the hit rates. The decrease in the drift rates are manifested as a decrease in accuracy and an increase in the reaction times and a spread in the reaction time distribution as a function of the test position.

If OI is due to the decrease in attention over the course of the test session, the boundary separation parameter should decrease as a function of the test position. The attention hypothesis suggests that at the beginning of the test, participants are more alert and spend more time on their recognition judgment and as a result more accurate. This hypothesis predicts a wider boundary separation at the beginning of the test and predicts the boundary separation to be narrower as a function of the test position. As a result, the attention hypothesis is supported if we observe a decrease in the response accuracy coupled with a decrease in the reaction time as the test position increases.

The SBME and OI

There were three goals of this research. The first one was to investigate the interaction between the SBME and OI as REM proposes a unified explanation for these effects. The SBME is explained by a reduction in the interference as a function of strengthening items

at study and OI is explained by an increase in the interference in memory as a function of encoding at retrieval. Thus, these two empirical findings can be explained by the interference that is caused by other items in memory. In the first experiment, the interaction between SBME and OI was investigated by administering a pure-study list paradigm in which participants studied either strong or weak words. REM predicts an interaction between the SBME and OI because the false alarm rate is higher after studying a weak list and as a result, more traces are incorrectly updated after endorsing a foil. Then, the likelihood of a match between a test item and traces in the memory will decrease more rapidly for weak test lists. This would result in a faster drop in the hit rates and a drop in the false alarm rates as a function of test position. In Experiment 1, the hit rates and the false alarm rates were analyzed as a function of list strength and the test position and both REM and the DM were fit to the data to evaluate the item-noise models, the criterion shift account and the attention hypothesis.

The second goal was to understand the mechanisms that cause SBME after studying a list of mixed strength items. As reviewed earlier, Starns et al. (2010; 2012) suggested that when participants were informed on the strength of the targets on the test list, they set different criteria across strong and weak test lists. In order to test whether informing participants changes the criteria placement at the beginning of the test, a mixed-study list paradigm was administered either by informing participants at the beginning of the test or not. Related to the second goal, we also investigated whether item-noise models predict differentiation at retrieval which would explain the SBME in a mixed-study list paradigm. Differentiation at retrieval was investigated by analyzing OI at different test list strength conditions after participants studied a list of words with mixed strength.

The final goal of this project was to differentiate between the item-noise and the attention hypothesis of OI using the DM. As the two explanations have different predictions in the DM, the parameters obtained from the best matching model were compared across three experiments to evaluate the most likely explanation of OI.

Experiment 1

In order to provide converging evidence for item-noise models, we investigated the effect of strengthening memory during study on OI. In this experiment, item strength is manipulated *across* lists by presenting pure strength study lists. As mentioned previously, REM predicts the SBME such that the hit rates increase and the false alarm rates decrease for the strongly encoded items. These predictions are based on the mechanism in REM that the strongly encoded items interfere less with other targets and the foils at test. REM also predicts OI due to encoding at retrieval. Accuracy will decrease as a function of test position as a result of both updating and storing new traces during test. More specifically, previous research showed a decrease mainly in the hit rates and relatively flat false alarm rates across test positions (Criss et al., 2011). REM also predicts an interaction between SBME and OI due to the higher false alarm rates in the weak lists. When the false alarm rates are greater, traces in memory will be updated incorrectly more often, and as a result, the match between the subsequent test items and memory will decrease. These predictions were also tested by fitting REM to accuracy data and will be discussed in detail in the results and discussions section. Further, the DM was fit to the reaction time distributions. Differentiation predicts a higher absolute value of the drift rates for both targets and foils in the strong list. The item-noise hypothesis of OI predicts a decrease in the drift rates of the targets across test position and the attention hypothesis of OI predicts a decrease in the boundary separation as a function of test position.

Methods

Participants

Thirty-four undergraduate students from the Syracuse University Research Participation Pool took part in the experiment. Twenty five of the participants were female. Six participants who had overall low accuracy ($d' < 0.5$) were excluded from the subsequent analyses.

Materials

The word pool consisted of nouns selected from MRC Psycholinguistics Database with a range of Kucera and Francis (1967) written frequency between 4 and 400 ($M = 39.77$), and number of letters between 4 and 8 (Coltheart, 1981). 2929 words made up the word pool when multiple forms of the same word (e.g. CHILD and CHILDREN) were excluded.

Procedure and Design

Participants completed 12 blocks of study and test in two sessions with 6 blocks in each session. The two sessions were scheduled on two consecutive days, so the sessions were approximately twenty-four hours apart. In each study block, participants were presented with 150 words and a levels-of-processing task was administered to manipulate the encoding strength (Craik & Lockhart, 1972). In each session, half of the blocks (3) were strong and the other half (3) were weak (Figure 4). The strong and weak blocks were randomly ordered for each participant. For the strong study lists, participants were asked to make a semantic judgment ("Does the word have a pleasant meaning?") and for the weak study lists, the judgment was orthographic ("Does the word contain the letter 'e'?"). The study trials were self-paced as the participants responded by pressing the 'z' or the '/?' keys on the keyboard and a 100 msec inter stimulus interval followed each response. The test list was constructed from 75 words that were randomly selected from the study

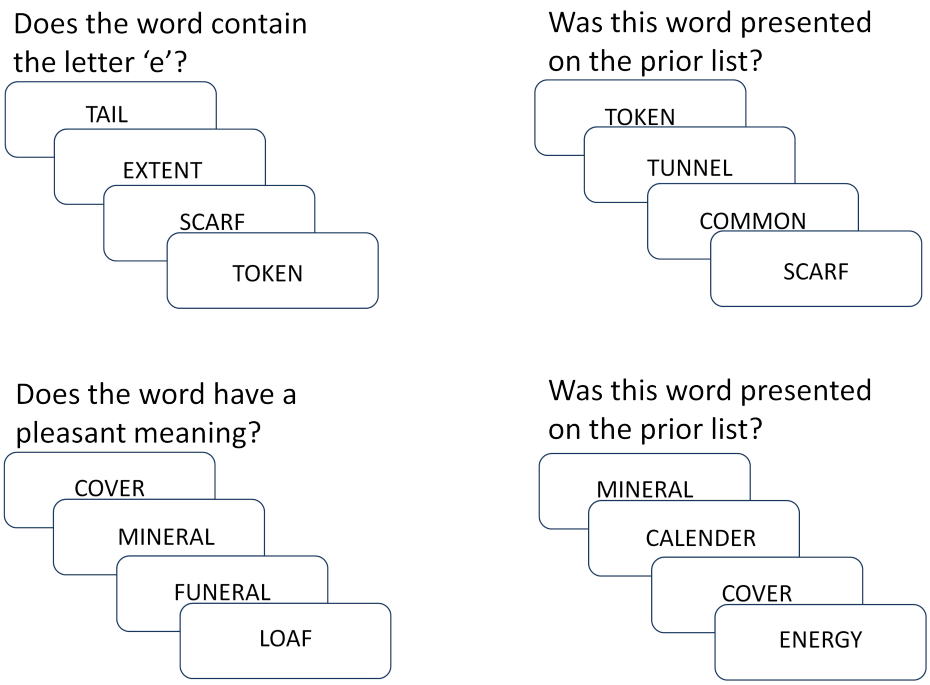

Figure 4: Design of Experiment 1. List strength was manipulated between study lists. The first column illustrates the study list and the second column illustrates the test list. First row is the weak study-test blocks and the second row is the strong study-test blocks.

list and 75 new words. For each test trial, participants were asked to make an 'old/new' recognition judgment by pressing 'z' for 'old' response and '/?' for 'new' response. The experiment was a 2 (Strength) × 5 (Test Block) within-subjects design. The responses in each study-test block was pooled with regards to the strength condition resulting in 6 study-test blocks for each strength condition. Later, test positions were binned into 5 test blocks which contain 180 items (150×6=900, 900÷5=180) in each for each participant. On average half of these 180 items were targets and the other were foils.

Results and Discussions

Accuracy

To assess the effects of strength and test position on the hit rates and the false alarm rates, the linear mixed effects model analysis was used with the `lme4` (Bates, Maechler, & Bolker,

2010) package in the `R environment` (R Development Core Team, 2011). In the mixed effects model, there are *fixed effects* which explain the variance in the dependent variable according to the experimental design and *random effects* which explain the between-subjects variance in the fixed effects. Random effects are normally distributed with a mean of 0 and the variance is estimated from the data. Variance (or standard deviation) of the random effect is the parameter of interest for measuring the between-subjects variability for the fixed effects. In other words, the estimated variance value shows how much the regression coefficients vary across individuals. More importantly, because the variance-covariance matrix is estimated for the random effects, a mixed effects model does not assume independence of variance across random effects and, as a result, provides an alternative to the sphericity (heterogeneity of covariance) assumption used in the repeated measures ANOVA. In the current analysis, a mixed effects model was fit separately to the hit rates and the false alarm rates with respect to inter-individual differences. The fixed effects correspond to the difference in the dependent variable at the group level associated with the experimental manipulations, whereas the random effects correspond to the variability in the fixed effects across 28 participants.

Hit Rates In order to assess the need for the mixed effects modelling, the intraclass correlation (ρ) was calculated for hit rates. The intraclass correlation is the ratio of between subjects variance to the total variance in hit rates [2]. The intraclass correlation obtained from the mixed effects model for the hit rates was 0.46 which suggests that 46% of the total variance in the data can be explained by individual differences. This indicates that the participants should be included as a second-level unit which accounts for the individual deviation from the experimental effect (Tabachnick & Fidell, 2006).

The model fits were progressed from the simplest model to the most complex model

[2]The intraclass correlation is calculated from an intercept-only model (null model) which is a model without any predictors. The intercept is added as both a fixed and a random variable, so each individual's deviation from the grand mean can be calculated. In order to obtain the intraclass correlation, variance of the random effect is divided by the total variance (sum of residuals and variance of the random effect).

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-166.58	-155.68	3	86.29	
2	Intercept, Strength	Intercept	-341.65	-327.11	4	174.82	M2-M1=177.07*
2a	Intercept, Test Position	Intercept	-241.78	-227.24	4	124.89	M2a-M1=77.19*
3	Intercept, Strength	Intercept, Strength	-346.52	-324.72	6	179.26	M3-M2=8.87*
3a	Intercept, Test Position	Intercept, Test Position	-238.48	-216.46	6	125.238	M3a-M2a=0.70
4	Intercept, Strength, Test Position	Intercept, Strength	-582.23	-556.79	7	298.12	M4-M3=237.71*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-588.40	-552.05	10	304.20	M5-M4=12.171*
6	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position	-614.08	-574.09	11	318.04	M6-M5=27.68*
7	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position, Interaction	-631.32	-576.79	15	330.66	M7-M6=25.24*

Table 1: Model comparisons for the hit rates (Experiment 1). The estimates are based on 280 data points. AIC is the Akaike Information Criterion, BIC is the Bayesian Information Criterion, log-lik is the log-Likelihood. The significant differences between the χ^2 values of each model pair at $\alpha = 0.05$ level is represented by * sign.

that explains the variance best. The simplest model was the intercept-only model which was used to assess the individual differences in the dependent variable (Model 1). Later, the list strength condition was added as a fixed effect which estimates the same strength effect for all participants (Model 2) and as a random effect (Model 3) which estimates a strength effect different across all participants. Table 1 summarizes all the models

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.70	0.03	21.01	0.17
Strength	0.14	0.02	6.54	0.08
Test Position	-0.07	0.006	-11.01	0.03
Strength × Test Position	0.025	0.006	3.88	0.03
Residual	-	-	-	0.05

Table 2: The best fitting mixed model parameters for the hit rates (Model 7) in Experiment 1. The strength effect has been treated as a factorial variable so the fixed effect estimate of the intercept is the estimated hit rate for the weak targets when the test position is equal to zero. t-values less than -2 or greater than 2 suggest a reliable fixed effect.

evaluated with their model fit values such as AIC, BIC, log-Likelihood and the degrees of freedom. The model fit values suggested that adding list strength as a predictor with a random effect improved the model by explaining the variance in the hit rates. Additionally, the test position was first added as a fixed effect only (Models 2a) and later with a random effect as well (Model 3a). In the later set of models, both the strength condition and the test position was evaluated with the fixed and the random effects (Model 4 and Model 5). The model comparison values suggested further improvement in the model fits as a result of these additions. Finally, the most complex model was the full model in which all the predictors (the list strength condition, test position and the interaction between the strength condition and test position) were added both as fixed (Model 6) and random effects (Model 7).

For the hit rates, Model 7 had the lowest AIC, BIC, and the highest log-Likelihood. The χ^2 difference test suggested that adding the strength effect, the test position and the interaction term as fixed and random effects increased the variance explained in the hit rates significantly. Table 2 presents the parameter estimates for the fixed and the random effects. The fixed effect estimates show that the hit rates for the strong targets were greater than the hit rates for the weak targets. The OI effect is also observed for the hit rates which decreased as the test position increased and the decrease in the hit rates as a

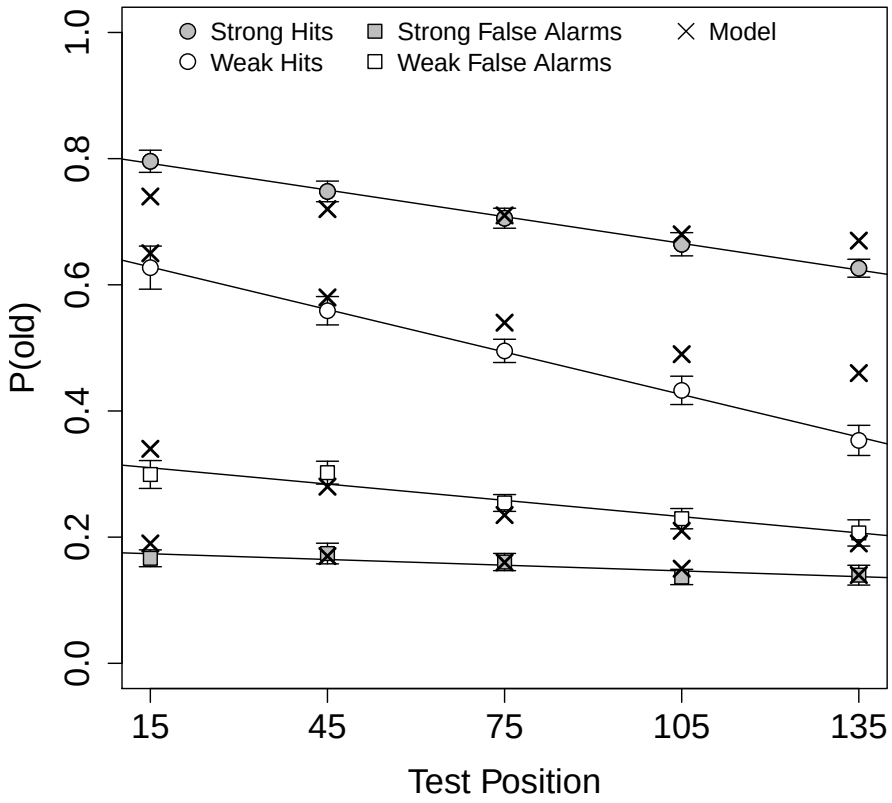

Figure 5: The hit rates and the false alarm rates as a function of the strength condition and the test position in Experiment 1. Lines represent the predictions from the best fitting linear mixed effects model (Model 7 for the hit rates and Model 6 for the false alarm rates). The points and the squares represent the data averaged across participants and × represents the predictions from REM. Error bars are within-subject 95% CI (Loftus & Masson, 1994).

function of test position was steeper for the weak targets. Figure 5 plots the hit rates averaged over participants as a function of test position separately for the strong and the weak targets. The model predictions were obtained from the fixed effects which were represented as lines. Figure 6 plots the hit rates as a function of the strength condition and test position for each individual with the model predictions at the individual level. The model predictions are plotted as lines. The figure suggests that all the participants showed the test position effect in the hit rates such that the hit rates of all participants decreased as a function of test position. The strength effect on the hit rates can also be seen from the

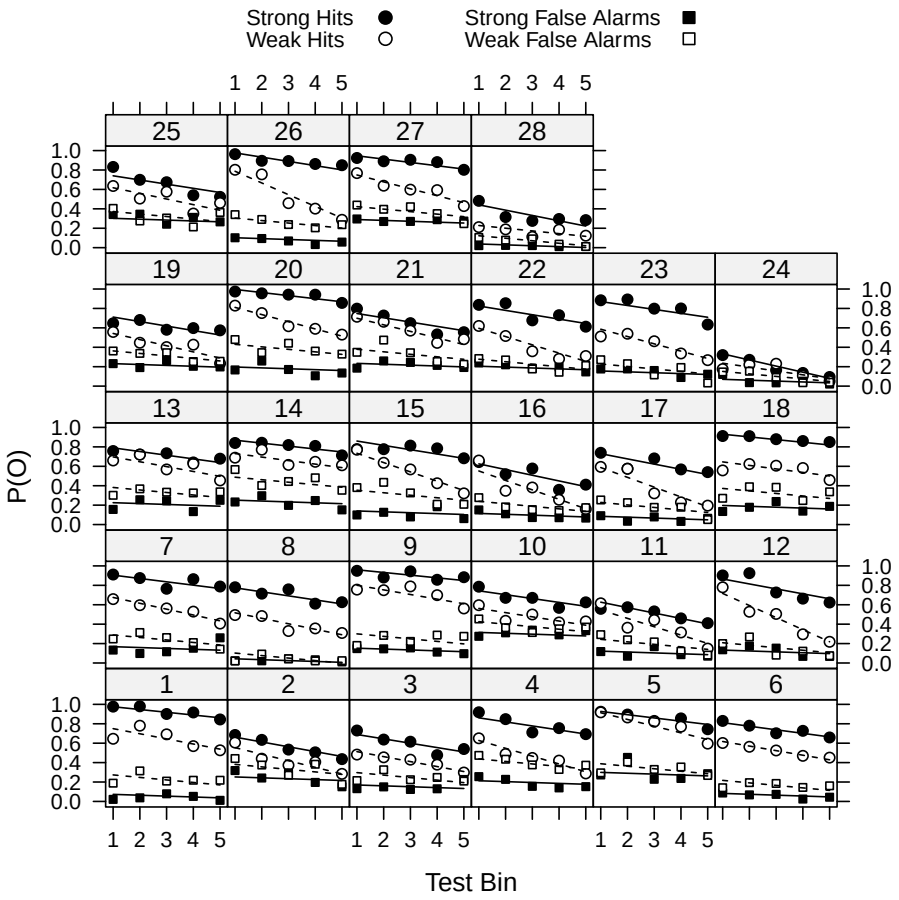

Figure 6: The hit rates and the false alarm rates as a function of the strength condition and the test position for each participant in Experiment 1. The data are represented by points and squares. The model predictions from the best fitting linear mixed effects model (Model 7 for the hit rates and Model 6 for the false alarm rates) are represented by the lines.

higher hit rates in strong targets for most of the participants. Additionally, the individual differences in the strength and test position effects can be observed visually. While some participants show a strong strength effect (e.g., Participants 23 and 18), others do not show such a strong effect (e.g., Participants 5 and 21). Similarly, while some participants show an interaction between the strength condition and test position (e.g., Participants 12 and 26) others do not (e.g., Participants 10 and 14).

False Alarm Rates The Intraclass correlation for the false alarm rates was 0.51, suggesting that 51% of the total variance in the false alarm rates could be explained by the

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-522.56	-511.66	3	264.28	
2	Intercept, Strength	Intercept	-658.53	-643.99	4	333.26	M2-M1=137.97*
2a	Intercept, Test Position	Intercept	-546.32	-531.78	4	277.16	M2a-M1=25.76*
3	Intercept, Strength	Intercept, Strength	-681.17	-659.36	6	346.58	M3-M2=26.64*
3a	Intercept, Test Position	Intercept, Test Position	-542.42	-520.62	6	277.21	M3a-M2a=0.10
4	Intercept, Strength, Test Position	Intercept, Strength	-740.00	-714.55	7	377.00	M4-M3=60.83*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-737.14	-700.19	10	378.57	M5-M4=3.14
6	Intercept, Strength, Test Position, Interaction	Intercept, Strength	-754.72	-725.64	8	385.36	M6-M4=16.72*
7	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Interaction	-750.36	-710.38	11	386.18	M7-M6=1.64

Table 3: Model comparisons for the false alarm rates (Experiment 1). The estimates are based on 280 data points. AIC is the Akaike Information Criterion, BIC is the Bayesian Information Criterion, log-lik is the log-Likelihood. The significant differences between the χ^2 values of each model pair at $\alpha = 0.05$ level is represented by * sign.

individual differences and showed a necessity to use the linear mixed effects model. Similar to the model fitting procedure applied to the hit rates, first the strength condition was added as a fixed effect (Model 2) and later also as a random effect (Model 3). Then, the test position was added separately first, as a fixed (Model 2a) and a random (Model 3a) effect, later combined with the strength condition effect (Model 4 and 5). Finally, the

interaction between the strength condition and the test position was tested through Models 6 and 7. Table 3 presents the fit values for the models that were evaluated to explain the false alarm rates.

The best fitting model (Model 6) shows a reliable effect of the strength condition on the false alarm rates such that the weak foils had higher false alarm rates compared with the strong foils (see Table 4 for parameter estimates). A reliable effect of test position shows that the false alarms rates decreased towards the end of the test list. Additionally, this decrease was more prominent for the weak foils. Model 6 also suggests that the strength effect varied across participants. The variation in the strength effect on the false alarms can be seen in Figure 6. However, relaxing the constraint on the test position effect (Model 5) did not improve the model fit significantly. This finding suggests that the effect of test position on the false alarms did not vary significantly across participants. Similarly, adding the interaction term as a random effect (Model 7) did not explain the variance significantly better than Model 6.

Figure 5 plots the average false alarm rates as a function of test position separately for the weak and the strong foils. The best fitting linear mixed effect model predictions (predictions from Model 6) are presented as lines. When the hit rates and false alarm rates are examined simultaneously, the SBME can be observed from the lower false alarms and the higher hit rates for the strong lists. In addition to the SBME, a negative effect of the test position can be observed for both hit rates and false alarm rates. This effect was more prominent for the weak items both in the hit rates and the false alarm rates.

REM All of these findings are predicted by an item-noise model, REM. One powerful aspect of REM is that it is a process model rather than a statistical model like the mixed effects model analysis which was previously used mainly to describe the data. Instead, REM simulates the mechanisms behind the item recognition and produces hit rates and false alarm rates based on a theoretical account. Additionally, REM predicts hit rates and

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.34	0.022	15.63	0.10
Strength	-0.15	0.017	-8.77	0.06
Test Position	-0.03	0.002	-9.04	-
Strength × Test Position	0.017	0.004	4.15	-
Residual	-	-	-	0.05

Table 4: The best fitting mixed model (Model 6) parameters for the false alarm rates Experiment 1. The strength effect has been treated as a factorial variable so the fixed effect estimate of the intercept is the estimated false alarm rate for the weak targets when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

false alarm rates simultaneously from the specified parameters in contrary to the statistical models that were fit to the hit rates and the false alarm rates separately. For the simulations in this project, the Criss et al. (2011) version of REM was used. Figure 5 presents the predicted hit rates and false alarm rates from REM for the data averaged over participants. The model fits were obtained by simulating 100 participants and averaging data from the simulated subjects. Only the u parameter was allowed to vary across study and test conditions, as u is the parameter that adjusts the encoding strength. u for the deep encoding task was 0.36 and u for the shallow encoding task was 0.21. The greater rate of decrease in the false alarm rates as a function of test position requires an increase in u for encoding during test. In the current simulation, u was fixed across strength conditions during test to 0.46. Having a greater u at test is also consistent with the testing effect which refers to an increase in accuracy after being tested on the previously studied items compared to restudying them (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). The *criterion* parameter was set to 0.75 and fixed across test lists. Other parameters of REM were also fixed and did not differ from their conventional values ($n = 20$, $g = 0.35$ and $c = 0.7$). For a summary of all of the parameter values used in this project see Table 5.

These findings suggest that with slight and meaningful changes in the parameters, REM can capture the qualitative pattern in the data including the interaction observed

Parameter	Value	Description
n	20	Vector length of the items
g	0.35	Feature frequency (geometric distribution parameter)
c	0.7	Probability of correctly copying a feature
u_{strong}	0.36	Probability of storing a feature of a strong item
u_{weak}	0.21	Probability of storing a feature of a weak item
u_{Test}	0.46	Probability of storing a feature of the best matching trace at test
$criterion$		Threshold for endorsing an item
Exp 1 and 2	0.75	
Exp 3: Strong	0.80	
Exp 3: Weak	0.70	

Table 5: The parameter values used in the REM simulations. In the REM simulations of Experiment 1-3, same parameters were used except the criterion parameter. The criterion parameter varied in Experiment 3 across test list strength conditions as the participants were told about the test list condition and thus let to adapt a different criterion for those conditions.

both in the hit rates and the false alarm rates. More importantly, REM suggests that a mechanism which is based on only the interference from other items in memory, can provide a unified explanation for the two empirical findings: the SBME and OI.

Reaction Time and the Diffusion Model Analysis

To further test whether the decrease in accuracy towards the end of the test list is due to item-noise or attention, the reaction time data was analyzed as a function of the strength condition and test position. In the experiment, each participant was tested on two conditions: the test lists following a strong study list and a weak study list. In the reaction time analysis, the test positions were binned into 5 blocks as in the accuracy analysis and the test item type was added as a condition with two levels (targets and foils). As a result of this factorial analysis, there were 20 (2 [Strength] $\times$ 5 [Test Block] $\times$ 2 [Item Type]) conditions for each participant. The reaction time distributions for correct and error responses of these 20 conditions was analyzed with the DM.

Reaction Time In the following DM fits, the responses that have a reaction time below 0.3 sec or above 3 sec were treated as outliers and excluded from the analysis consistent

with other applications to recognition memory (Criss, 2010; Ratcliff, Thapar, & McKoon, 2004; Ratcliff & Tuerlinckx, 2002; Starns et al., 2012). The reaction time distributions of 20 conditions are illustrated in quantile plots as a function of accuracy for each strength condition and test block condition separately for the targets and the foils and for correct and incorrect responses (see Figure 7). The vertical points are the 0.10, 0.30, 0.50, 0.70 and 0.90 quantiles of the reaction times for each condition. For example, the lowest point represents the fastest 10% of the responses which suggests that 10% of the responses in that condition have a reaction time faster than that value. The next fastest 20% of the responses would be between 0.1 and 0.3 quantiles and so on. Similarly, the highest point in the same condition represents the slowest 10% of the responses such that 10% of the responses have a reaction time slower than that value. The point in the middle is the median value which means that the 50% of the responses are slower and the other 50% of the responses are faster than the reaction time in the middle. When the spread of the points along the y-axis are examined for a given condition, the shape of the reaction time distributions can be easily interpreted. For example, the spread below the median is narrower compared to the spread above the median which suggests that the reaction times are positively skewed. Thus, these plots are useful for observing the changes in the shape of the reaction time distributions as a function of the accuracy of each condition. The figure also plots the 'old' and the 'new' responses given to each condition separately. The circles represent the 'old' responses and triangles represent the 'new' responses. The darkest symbols represent the earliest test position bin (average test position is 15) and the lightest symbols represent the latest test position bin (average test position is 135).

Figure 7 shows an effect of the strength condition on the reaction time quantiles such that the spread in the quantiles along the y-axis changes across the weak and the strong test lists. The effect of the strength condition on accuracy can be observed from the x-axis. For example, the weak targets have a lower proportion of 'old' responses compared to the strong targets while the 'old' response proportion for weak foils are higher than that

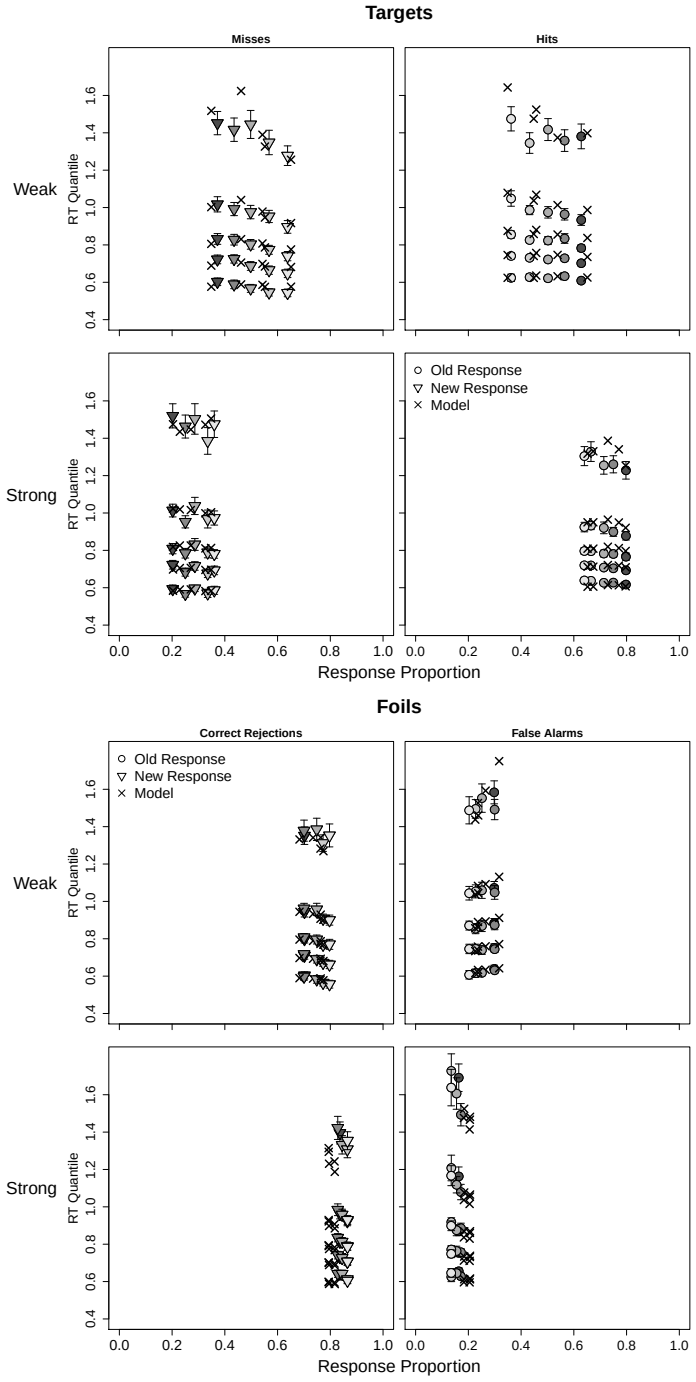

Figure 7: The reaction time distributions for Experiment 1 are presented as quantiles as a function of accuracy. Each vertical plot corresponds to one condition which is a combination of the test position, the strength condition, the item type and the response. Test position is represented by the shades of grey: The darkest point is the first test position bin and the lightest point is the last test position bin. Error bars are the standard error mean for each reaction time quantile. $\times$ represents the model predictions from DM #2 (group data).

of the strong foils. In addition to the strength effect, the increase in the test positions also changes the shape of the reaction time distributions of both weak and strong test lists. All of these changes in the shape of the reaction time distributions coupled with the item type and accuracy are due to decision and memory processes. However, these processes are entangled and visual investigations would not be enough to derive conclusions about the attention, the criterion shift or the differentiation accounts. To quantify the effects on decision and memory processes, the DM was fit to the reaction time and accuracy data. In the next section, DM fits will be presented.

Diffusion Model Analysis Models were fit to the group and the individual data both. The "Vincentizing" procedure was applied in order to average the reaction time distributions for the group data (Ratcliff, 1979; Vincent, 1912). First, the reaction time quantiles were calculated for each individual participant and, later the reaction time quantiles were averaged across participants. The pooled response frequencies were binned with respect to those averaged reaction time quantiles. The models were fit with the Diffusion Model Analysis Toolbox in MATLAB (Vandekerckhove & Tuerlinckx, 2007, 2008) by minimizing the χ^2 statistic which computes the deviance between the observed and the predicted values.

Two models were fit to the group data. The first model (DM #1) was fit to test SBME and the item-noise account of OI by varying the drift rates (v) across all 20 conditions which correspond to the combination of 5 test position bins, 2 strength conditions and 2 item types. Thus, this model is consistent with the item-noise account of the SBME and OI in accordance to this account, both of these effects are based on memory, thus the drift rate parameter of the DM. In addition to the drift rates, the starting point parameter (z) was allowed to vary across strength condition as previous studies of SBME in DM showed an effect of the strength condition on this parameter (Criss, 2010; Starns et al., 2012). In summary, there were 27 free model parameters (20

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)	χ^2 (Ind.)
1	1758	27	228083	0.00	228321	0.00	307
2	1675	29	228028	1.00	228283	1.00	299

Table 6: The diffusion model comparisons in Experiment 1. χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion with finite sample correction, w_i(AICc) is the Akaike weights which represents the probability of the model being the best model, BIC is the Bayesian Information Criterion, w_i(BIC) is the Schwarz weights which represent the probability of the model being the best model according to BIC values. The last column is the average χ^2 value from the fits to individual participants.

from v, 2 from z, a, T_{er}, s_T, s_z, η) and the degrees of freedom in the data was 220 [3].

The second model (DM #2) tested the attention hypothesis of OI by allowing the boundary separation parameter (a) vary across test positions in addition to the drift rate parameter. In this model, the response bias (i.e., the relative starting point parameter, z/a) was fixed across test positions as in DM #1. However, in order to fix the response bias, the starting point parameter (z) is also required to vary proportionally to the boundary separation parameter (a). In the model, the starting point is defined as a value between 0 and a. The zero point of the boundary separation represents the 'new' response boundary and a represents the 'old' response boundary. If the boundary separation parameter is allowed to vary while the starting point parameter is fixed, the change in a will mimic a change in the response bias. For example, as the boundary separation value increases, the 'old' boundary shifts farther from the starting point and the place of the 'new' boundary is fixed (the zero point). Thus, this pattern will mimic a shift in the response bias towards the 'new' response boundary. In order to control for this confound, the response bias (z/a) was fixed in the model by constraining z to vary relative to a. That response bias value was obtained from the first model in which both the boundary separation and the starting point parameters were fixed across test positions. However, this

[3]For each condition, there were 12 response frequencies as there were 5 reaction time quantiles (i.e. 6 response frequency bins) for each response ('old' and 'new'). As a result, there were 11 degrees of freedom in each condition since one degree of freedom is lost due to the restriction that all the response frequencies must add up to the total number of responses. Since there were 20 conditions, degrees of freedom in the data was 220 in all of the DM fits that will be discussed in this project.

bias value was different across the strength conditions as the starting point parameter was let to vary across list strength. As a result, two different bias values were specified in the second model, one for the weak list and another for the strong list. In summary, the degrees of freedom in this model increased to 29 (20 from v, 5 from a, T_{er}, s_T, s_z, η, no parameters for z as it varied proportionally to a with a fixed bias value). Additionally, a comparison of the two models allows us to test the effect of boundary separation directly because the relative position of the starting points are fixed across the two models.

Model fit statistics are presented in Table 6. The table lists the Akaike Information Criterion (AICc), the Bayesian Information Criterion (BIC) and their weights for model selection (Burnham & Anderson, 2004; Kass & Raftery, 1995; Wagenmakers & Farrell, 2004) for the group data. The AICc and the BIC was highest for the first model in which the drift rates varied across all conditions and the starting point parameter varied only across list strength. Thus, the model fit statistics suggest that the model could be improved by varying the boundary separation across test positions in addition to drift rates. Both BIC and AICc favor DM #2. For further comparison, both models were also fit to individual data and the results from the individual fits show that the data from fifteen participants were best fit by DM #2, the data from six participants were best fit by DM #1 and for the rest of the participants ($N = 7$), AICc and BIC preferred different models.

The parameters of the two models from the group data fit are presented in Table 7. The average parameters from the individual data fits are presented in Table A.13. The drift rate parameters obtained from DM #2 provided evidence for the SBME. The drift rates from the individual fits were analyzed by a 2 (Strength) $\times$ 5 (Test Block) repeated measures ANOVA separately for targets and foils. The strength effect was significant for both targets, $F(1, 260) = 22.64$, $p < .001$ and foils, $F(1, 260) = 26.78$, $p < .001$. The mean drift rates of the strong targets were greater ($M = 0.187$, $SD = 0.16$) than the mean drift rates of the weak targets ($M = 0.037$, $SD = 0.13$) and the mean drift rates of the strong foils ($M = -0.225$, $SD = 0.10$) were greater in absolute value than the mean drift rates of

Group Parameter Values (DM #1)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.144	0.308	−0.147	-0.262	0.186	
2	0.086	0.265	−0.166	-0.262	0.186	
3	−0.008	0.194	−0.200	-0.278	0.186	
4	−0.028	0.172	−0.203	-0.312	0.186	
5	−0.074	0.158	−0.228	-0.301	0.186	
Fixed Parameters	z/a_W 0.45	z/a_S 0.48	η 0.237	s_z 0.165	T_{er} 0.587	s_T 0.301

Group Parameter Values (DM #2)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.163	0.291	−0.176	-0.254	0.183	
2	0.074	0.251	−0.136	-0.252	0.185	
3	−0.010	0.208	−0.202	-0.285	0.183	
4	−0.016	0.167	−0.204	-0.296	0.176	
5	−0.110	0.150	−0.216	-0.264	0.175	
Fixed Parameters	z/a_W 0.45	z/a_S 0.48	η 0.248	s_z 0.157	T_{er} 0.587	s_T 0.292

Table 7: The group parameter values of DM from Experiment 1. In the first model, the boundary separation was fixed across the test blocks.

the weak foils ($M = -0.143$, $SD = 0.09$). Figure 8 plots the box-and-whisker diagram of the drift rates as a function of strength. These findings support the differentiation account for the SBME as the foils were less confusable in the strong test lists and as a result, evidence accumulated faster towards the 'new' boundary. Similarly, the drift rates of the targets decreased significantly as a function of test block, $F(4, 260) = 3.8$, $p < .01$; providing evidence for the item-noise account of OI. The mean drift rate at the first test block ($M = 0.205$, $SD = 0.15$) was significantly greater than the mean drift rate at the last test block ($M = 0.02$, $SD = 0.16$), $t(55) = 10.58$, $p < .001$. The drift rates of the foils did not differ significantly as a function of test block. Likewise, the interaction between the

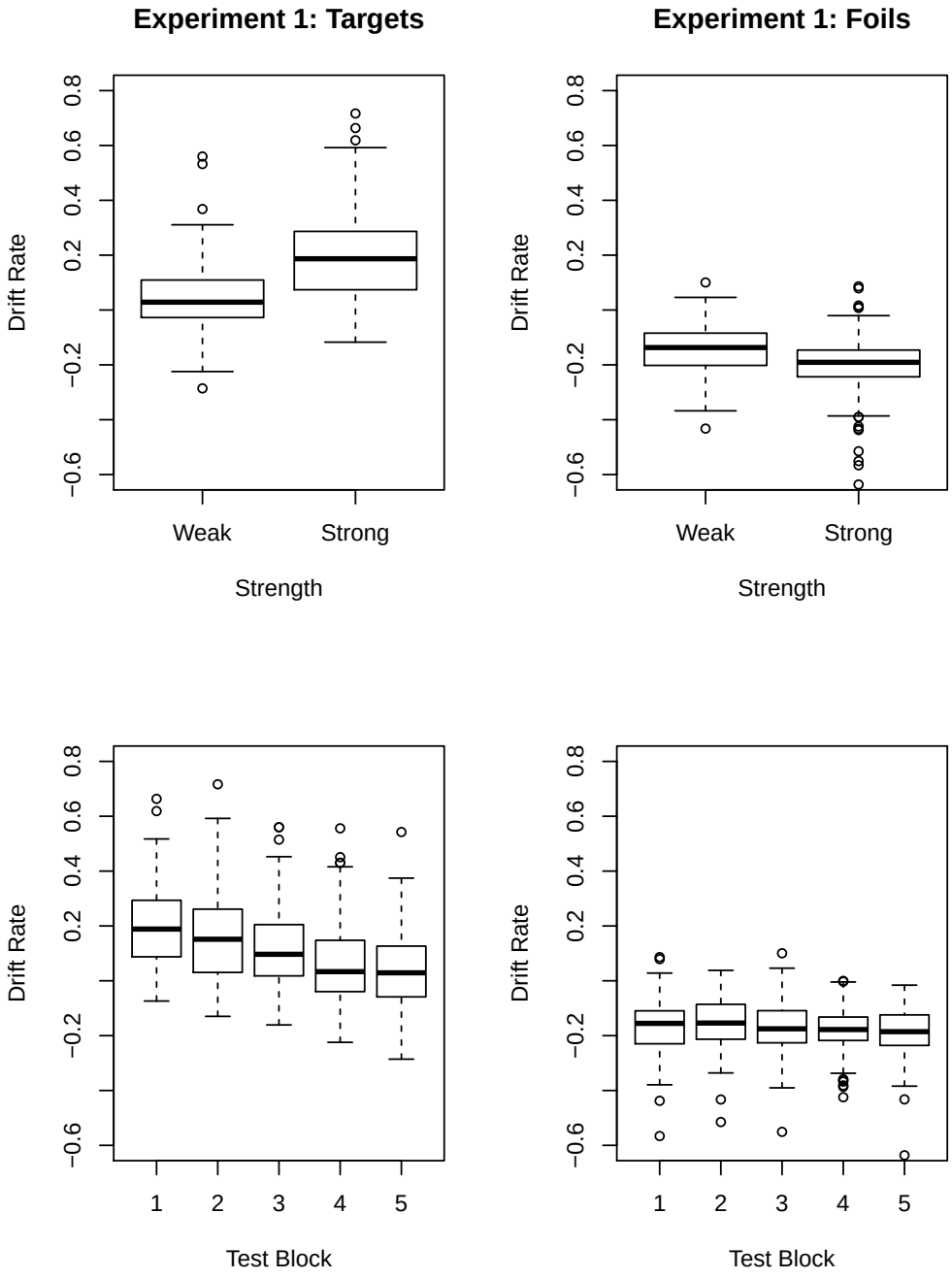

Figure 8: Box-plots for the drift rate parameter estimates obtained from DM#2 that was fit to individual data. The first row plots the drift rate estimates from each individual as a function of the strength condition for targets and foils. The second row plots the drift rate estimates as a function test block.

strength and the test block did not reach significance.

A response bias (z/a) value of 0.5 represents an unbiased response criterion which

means the evidence accumulation starts from the middle point of the distance between the boundaries. A response bias value lower than 0.5 represents a more stringent criterion as more evidence is required to reach the 'old' boundary because the starting point is closer to the 'new' boundary. The relative starting point parameter estimates (estimated response bias) obtained from DM#1 for each individual was compared across strength conditions. The mean relative starting point parameter value was higher for the strong items ($M = 0.50$, $SD = 0.08$) than for the weak items ($M = 0.45$, $SD = 0.08$), $t(27) = -5.07$, $p < .001$ (see Figure 9 for box-plots). This is the opposite of what criterion shift account would predict for the SBME because this pattern suggests that the starting point is closer to the 'old' boundary for strong items (see also Criss, 2010; Starns et al., 2012). Thus, these parameter values show that strong items start closer to the 'old' boundary and therefore reach the boundary faster and more often for a given accuracy level. However, Starns et al. (2012) argue that the criterion that shifts to explain the SBME is the drift criterion and does not necessarily manifest itself as the starting point parameter. However, the drift criterion can not be measured independent of the drift rates. Thus, these data do not provide a direct evidence for the criterion shift account.

The boundary separation (a) was the only parameter that was specified differently in the two competing models. In DM #2, the boundary separation parameter value was allowed to vary across test blocks in order to test the attention hypothesis of OI. In order to eliminate the confounding effect of fixing the starting point parameter across test blocks, the starting point was allowed to vary relative to the boundary separation with a restriction of a specific bias value obtained from DM #1. The boundary separation parameter decreased towards the end of the test list, especially the last two test blocks which correspond to the last 60 test trials (Figure 9). A one way ANOVA was conducted to test the effect of test block on the boundary separation parameter and the results revealed a significant effect, $F(4, 108) = 15.29$, $p < .001$. The average boundary separation estimates did not differ significantly across the first three test blocks but decreased

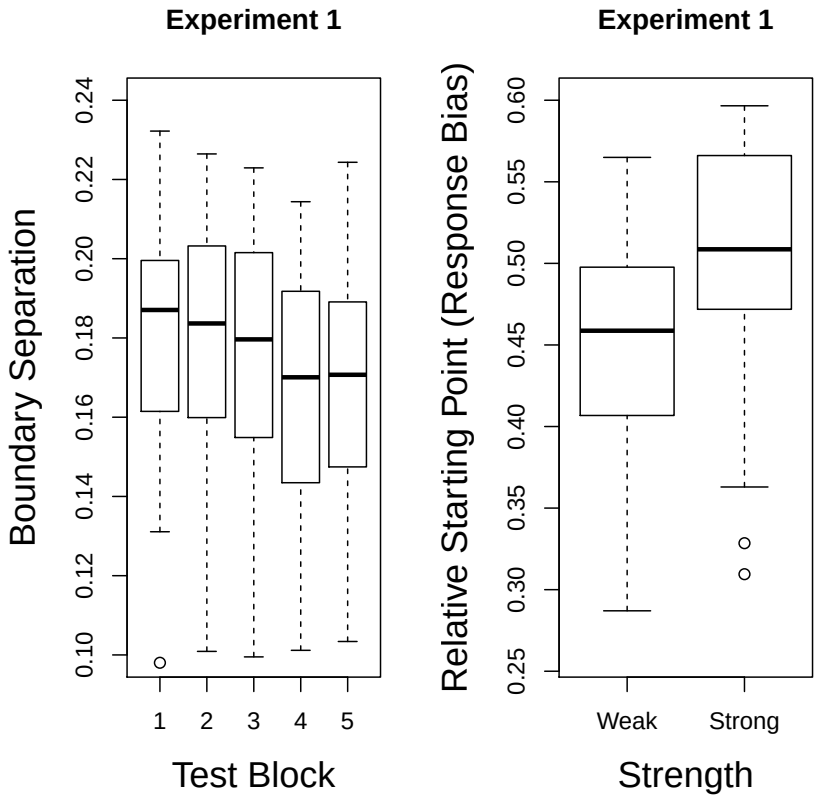

Figure 9: On the left, box-plots for the boundary separation parameter estimates obtained from DM#2 that was fit to individual data. On the right, box-plots for the response bias estimate from DM#1 that was fit to each individual.

significantly from third block ($M = 0.176$, $SD = 0.03$) to the fourth block ($M = 0.170$, $SD = 0.03$), $t(27) = 2.70$, $p = 0.01$. These findings support the attention hypothesis for OI, which states that the accuracy decreases as a function of test position due to a decrease in attention. However, we should mention that, this effect is observed in addition to the decrease in the drift rates across test blocks. The results from the accuracy analysis and the quantile plots show that accuracy decreased monotonically as a function of test position, yet the boundary separation parameter shows a non-monotonic pattern. Thus, the OI observed in this experiment is best explained by the decrease in the drift rates which show a similar monotonic decrease as in the accuracy analysis, rather than a decrease in attention across the test list.

To summarize, the results from the reaction time and the DM analysis suggest that both memory and decision processes contribute to SBME and OI. Higher absolute values of the drift rate parameters for the strong items suggest that deep encoding during study results in more memory evidence favoring the 'old' response for the targets and the 'new' response for the foils. According to item-noise models, this difference in the memory evidence is due to interference from other items in memory during a recognition task. The difference between the starting point parameter values across strength conditions also suggests that different decision criteria are used in recognition memory for different strength conditions. However, these differences in the decision criteria are not consistent with what criterion shift account proposes for the SBME.

The parameter values regarding OI also provides evidence for the item-noise models. As the test progressed, the drift rate parameter values decreased. This shows a decrease in the rate of the accumulation of evidence which can be observed as increasing reaction times and decreasing accuracy towards the end of the test list. Finally, the decrease in the boundary separation parameter at the end of the test provides evidence for the attention hypothesis as a contributing factor to the decrease in the accuracy observed at the very end of the test list.

Summary

The results from Experiment 1 provided converging evidence for the item-noise models by showing that strengthening memory during study has an effect on output interference. The results from the accuracy analysis and the DM fits showed the SBME such that the hit rates increased and the false alarm rates decreased for the strongly encoded items. The greater value of the drift rates for the strong targets further supported REM showing that the rate of memory evidence accumulation is faster for the strongly encoded items and the greater absolute value of the drift rates for the strong foils showed that the these foils were

less confusable with the strongly encoded items. Additionally, OI was also observed in this experiment such that the hit rates decreased as a function of the test position. The results from the DM fits showed that both memory evidence and attention play a role in the decrease in the accuracy towards the end of the test list. Finally, the interaction between the SBME and OI was present in the results along with the predictions from REM. The results from the accuracy analysis showed that strengthening memory during study reduced the detrimental effect of test position. This interaction pattern could also be seen in the drift rate parameter values from the DM fits.

Experiment 2

Item-noise models have recently been challenged by Starns and colleagues (e.g. Starns et al., 2010, 2012) who showed that a SBME could be observed when item strength is manipulated *within* study lists. Although item-noise models predict an increase in the hit rates for strongly encoded items, these models do not predict any difference in the false alarm rates when the encoding conditions are constant. For example in REM, memory evidence for test items is assessed through the match between the test item and all of the traces of the studied items in memory. When strength is manipulated only at test, foil items from both test conditions are matched to the same set of strong and weak traces resulting in approximately the same distribution of subjective memory strength. Thus, REM does not predict different false alarm rates between the two strength conditions based the memory strength only. It would require changes in the criterion across the strength conditions in order to account for that SBME. Thus, Starns et al. argued that the SBME observed in a mixed-study paradigm could be considered evidence for the criterion shift account of SBME in a pure-list paradigm. In other words, REM predicts the pure-list SBME results from differentiation but the mixed-list SBME results from a criterion shift whereas Starns et al. (2010, 2012) predicts that both effects are due to a criterion shift. In this experiment, participants studied a mixed-strength list and were tested with either strong or weak targets to evaluate whether the SBME is due to the differentiation at retrieval or to the decision processes. In order to control for the possible criterion shift at the beginning of the test, participants were *not informed* on the strength of the targets

that they would be tested on, contrary to the Starns et al. paradigm. The accuracy data from this experiment will be discussed along with the predictions from REM with encoding at retrieval. Furthermore, the DM will be used to investigate the SBME and OI with the reaction time data.

Methods

Participants

Thirty-two Syracuse University undergraduates were recruited from the Research Participation Pool in exchange for course credit. Thirteen of the participants were female. Participants who had overall low accuracy level ($d' < 0.5$) were excluded from the subsequent analysis which resulted in a sample size of twenty six participants for the subsequent analysis.

Materials

The lists were constructed from the same word pool used in Experiment 1.

Procedures and Design

The procedures were very similar to Experiment 1 except that the words were strengthened within lists during study. Participants completed 12 blocks of study and test in two sessions with 6 blocks in each session. The two sessions were scheduled on two consecutive days, so that the sessions were approximately twenty-four hours apart. In each study block, the participants were presented with 150 words and were asked to make a semantic judgment ("Does the word have a pleasant meaning?") for 75 of the words and an orthographic judgment ("Does the word contain the letter 'e'?") for the other 75. The order of the encoding tasks was random and the participants responded by pressing the 'z'

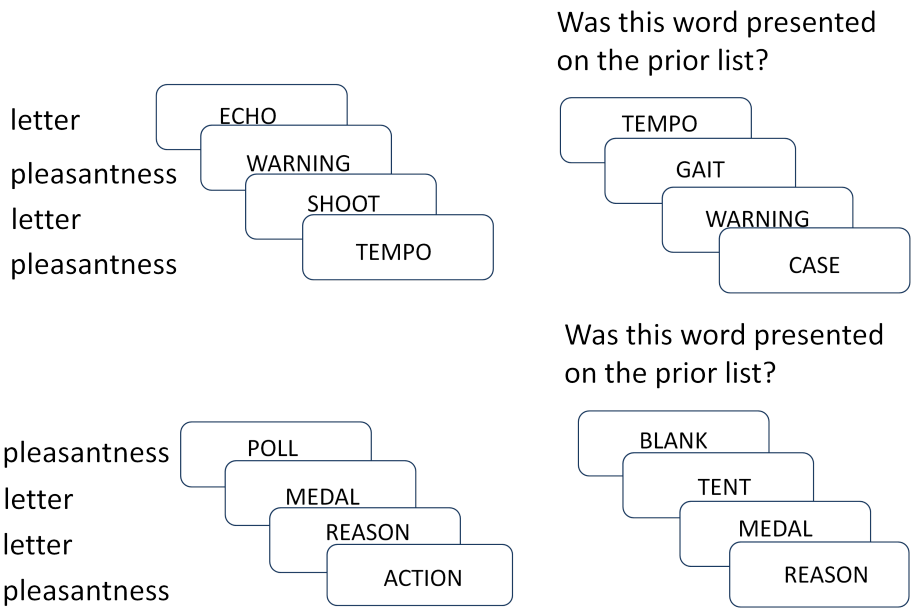

Figure 10: The design of Experiment 2. List strength was manipulated within study lists. The first column illustrates the study lists and the second column illustrates the test lists. First row is the strong condition in which the targets in the test list are the weakly studied items and the second row is the weak condition in which the targets are the strongly studied items.

or the '/?' keys on the keyboard. After the participants responded, there was a 100 msec inter stimulus interval before moving on to the next study trial. Strength was manipulated at test by presenting pure strength test lists (Figure 10). In each session, half of the blocks (3) were randomly selected, to test only the strongly encoded words along with 75 new words (strong test list). In the other half of the blocks (3), only the weakly encoded words were used to construct the test list with an additional 75 new words (weak test list). In each test trial, participants were asked to make an 'old/new' recognition judgment. The participants were not informed on which type of list (strong or weak) they would be tested. The experiment was a 2 (Strength) $\times$ 5 (Test Block) within-subjects design. Test positions were binned into 5 test blocks, and each test block contained 180 trials, where the test lists with the same strength condition were pooled for each participant.

Results and Discussions

Accuracy

The linear mixed effects model analysis was used to assess the effects of strength and test position on the hit rates and the false alarm rates. Initially, intraclass correlation was calculated in order to test the necessity for using the mixed effects model. Later, in order to select the model that best explains the hit rates and the false alarm rates, models with different complexity were fit to the data. The model fit statistics were presented and the model that explained the data with the minimum number of degrees of freedom was selected as the best model.

Hit Rates The intraclass correlation obtained from the mixed effects model for the hit rates was 0.27, which means that 27% of the total variance was due to individual differences. Thus, individuals were added as a second-level unit to explain the between-subject variance (Tabachnick & Fidell, 2006). Firstly, the strength condition was added as a fixed effect to test the effect at the group level (Model 2), and later, as a random effect which allows the effect of the strength condition vary across individuals (Model 3). Later, to test OI, the test position was added as a fixed effect (Model 2a), and then, as a random effect (Model 3a). The next models assessed the combined effects of the strength condition and the test position both as fixed and random variables (Models 4 and 5). The interaction between the strength condition and the test position was added to the model as a predictor with fixed effects (Models 6) and with random effects (Model 7). Finally, a reduced model (Model 8) was fit to further test the fixed effect of the interaction term. Table 8 summarizes all of the models evaluated with their model fit values for AIC, BIC, the log-Likelihood and the degrees of freedom.

The model fit statistics suggest that Model 8 is best at explaining the data with minimum complexity. From Table 8 shows Model 8 is a reduced model of Model 7. In

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-149.35	-138.67	3	77.675	
2	Intercept, Strength	Intercept	-337.53	-323.39	4	172.765	M2-M1=190.18*
2a	Intercept, Test Position	Intercept	-187.23	-172.99	4	97.615	M2a-M1=39.88*
3	Intercept, Strength	Intercept, Strength	-367.50	-346.13	6	189.75	M3-M2=33.97*
3a	Intercept, Test Position	Intercept, Test Position	-183.56	-162.20	6	97.78	M3a-M2a=0.33
4	Intercept, Strength, Test Position	Intercept, Strength	-526.29	-501.37	7	270.15	M4-M3=160.79*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-537.22	-501.62	10	278.61	M5-M4=16.93*
6	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position	-541.37	-502.20	11	281.68	M6-M5=6.14*
7	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position, Interaction	-566.76	-513.35	15	298.38	M7-M6=33.39*
8	Intercept, Strength, Test Position	Intercept, Strength, Test Position, Interaction	-566.17	-513.32	14	297.09	M7-M8=2.59

Table 8: The model comparisons for the hit rates (Experiment 2). The estimates are based on 260 data points. AIC is the Akaike Information Criterion, BIC is the Bayesian Information Criterion, log-lik is the log-Likelihood. The significant differences between the χ^2 of each model pair at $\alpha = 0.05$ level is represented by * sign.

Model 7, the effects of the strength condition, the test position and the interaction between the two were defined as both fixed and random effects. However, the fixed effect parameter for the interaction term was not reliably different from zero (0.013, $SE = 0.008$, $t = 1.62$).

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.62	0.02	26.14	0.17
Strength	0.23	0.02	9.43	0.15
Test Position	-0.04	0.004	-10.94	0.03
Strength × Test Position	-	-	-	0.03
Residual	-	-	-	0.05

Table 9: The best fitting mixed model parameters for the hit rates (Model 8) in Experiment 2. The strength effect has been treated as a factorial variable so the fixed effect estimate of the intercept is the estimated hit rate for the weak targets when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

As a result, a model without the interaction term as a fixed effect was fit to the data and later compared with the full model. AIC, BIC and the χ^2 difference test showed that removing the interaction term as a fixed effect did not harm the model fit. The resulting model parameter estimates are presented in Table 9.

The fixed effect estimates show that the strongly encoded targets were endorsed more often than the weakly encoded targets, and the hit rates decreased as a function of test position. Figure 11 plots the hit rates averaged across participants as a function of the strength condition and the test position. The lines are the predicted values from the fixed effects of the best fitting statistical model (Model 8). The random effect estimates in Model 8 show that the effects of the strength condition, the test position and the interaction between the two predictors reliably depend on the individual. However, the lack of a reliable fixed effect for the intercept term suggests that there was not a consistent effect of an interaction between the strength and the test position at the group level. Figure 12 plots each individual's data with model predictions at the individual level. As most of the participants show strength and test position effects, only a small subset of participants show an interaction between the two predictors. Additionally, some participants (e.g. participants 18 and 20) showed an inverted pattern such that the hit rates increasing as a function of test position for the weak targets instead of decreasing

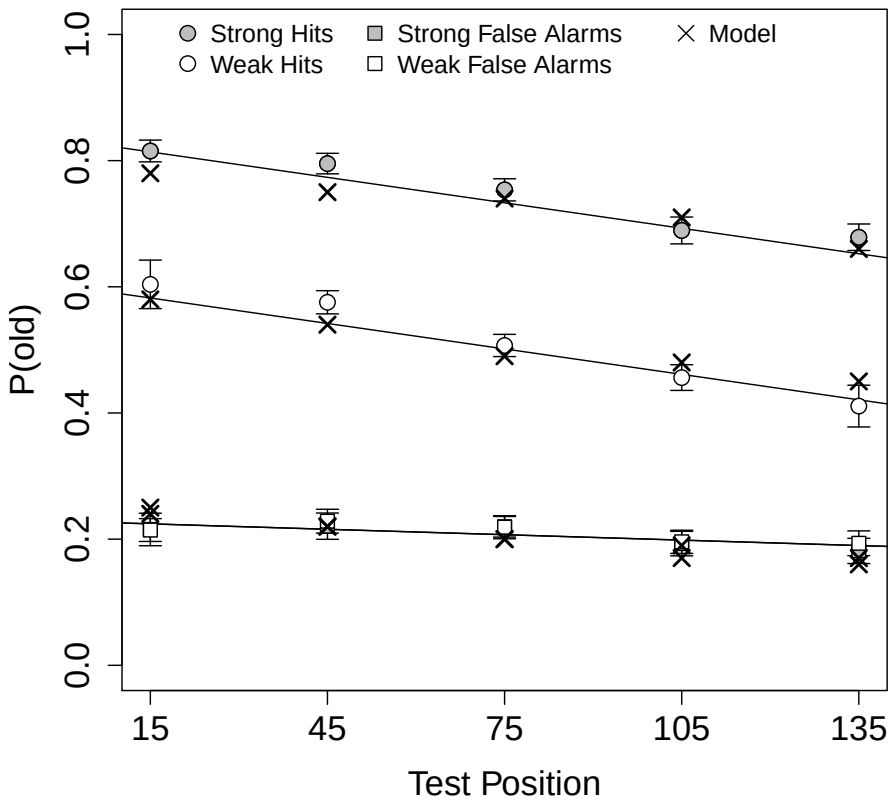

Figure 11: The hit rates and the false alarm rates as a function of the strength condition and the test position for Experiment 2. The lines represent the predicted hit rates and the predicted false alarm rates from the best fitting linear mixed effects model (Model 8 for the hit rates and Model 4 for the false alarm rates). The points and the squares represent the averaged hit rates and the averaged false alarm rates over participants. $\times$ represents the predicted hit rates and the predicted false alarm rates from REM. Error bars are within-subject 95% CI.

faster as the mean interaction term would suggest. This figure is consistent with Model 8 as the effect of the interaction term can not be generalized at the group level. As a result, when the individual differences in the interaction term are statistically controlled, the fixed effect becomes statistically unreliable. This shows that there is high variability in the SBME $\times$ OI interaction.

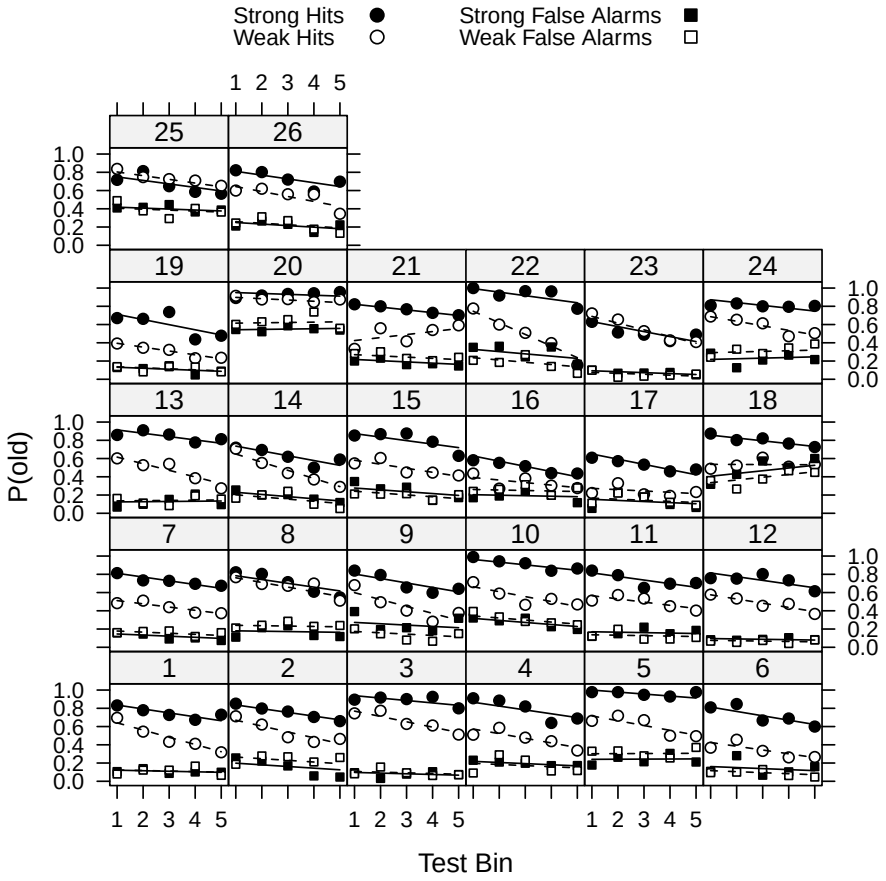

Figure 12: The hit rates and the false alarm rates as a function of the strength condition and the test position for each individual participant in Experiment 2. The data are represented by the points and the squares. The linear mixed effects model predictions are represented by the lines (Model 8 for the hit rates and Model 4 for the false alarm rates).

False Alarm Rates The intraclass correlation for the false alarm rates shows that 22% of the total variance in the false alarm rates can be explained by the individual differences. In order to assess the effects of the strength condition and the test position on the false alarm rates, multiple models were fit to the data. Table 10 presents the model comparisons. Initially, the strength condition was added as a fixed effect (Model 2) to the null model and the model comparison statistics suggested that this addition did not improve the model fit significantly. Later, the strength condition was added both as a fixed and a random effect (Model 3) and compared to Model 2. The improved model fit suggested that the effect of the strength condition differed at the individual level. A similar

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-604.88	-594.19	3	305.44	
2	Intercept, Strength	Intercept	-603.18	-588.93	4	305.59	M2-M1=0.30
2a	Intercept, Test Position	Intercept	-612.71	-598.47	4	310.36	M2a-M1=9.84*
3	Intercept, Strength	Intercept, Strength	-620.93	-599.56	6	316.46	M3-M2=21.75*
3a	Intercept, Test Position	Intercept, Test Position	-615.29	-593.93	6	313.65	M3a-M2a=6.58*
4	Intercept, Test Position	Intercept, Test Position, Strength	-641.40	-609.36	9	329.70	M4-M3a=469.52*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-639.51	-603.90	10	329.75	M5-M4=0.105

Table 10: The model comparisons for the false alarm rates (Experiment 2). The estimates are based on 260 data points. AIC is the Akaike Information Criterion, BIC is the Bayesian Information Criterion, log-lik is the log-Likelihood. The significant differences between the χ^2 of each model pair at $\alpha = 0.05$ level is represented by * sign.

procedure was followed to assess the effect of test position. Comparisons between Model 1 and Model 2a, as well as between Model 2a and Model 3a, showed that adding the test position as a fixed and a random effect explains the variance in the false alarm rates better. As a result, a model (Model 4) with only the test position as a fixed effect and both test position and the strength condition as a random effect was compared to Model 3a. A significant χ^2 difference test coupled with lower AIC and BIC values suggested that statistically controlling the individual differences in the strength effect best explained the variance in the data with the minimum degrees of freedom. Finally, to test the effect of the strength condition at the group level, the strength effect was added as a fixed effect to Model 4 yielding to Model 5. Neither AIC, BIC, nor the χ^2 difference test suggested a

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.23	0.02	10.07	0.12
Test Position	-0.008	0.004	-2.42	0.01
Strength	-	-	-	0.04
Residual	-	-	-	0.05

Table 11: The best fitting mixed model parameters for the false rates (Model 4) in Experiment 2. The intercept is the estimated hit rate when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

reliable effect of the strength condition on the false alarms since adding one more free parameter did not improve the model fit. This finding suggests that some participants had a higher false alarm rate and the others had a lower false alarm rate in the weak test lists compared to their false alarm rates in the strong test lists. Thus, the effect of the strength condition can not be generalized at the group level. Model 4 was chosen as the best model and the parameter estimates are displayed in Table 11. Although the fixed effect estimate for the test position was small, the low inter-individual variability in the test position effect suggests that the effect is reliable. To further test the significance of the test position fixed effect, 10000 Markov Chain Monte Carlo (MCMC) samples were generated from the posterior distribution of the parameters of Model 4. The mean fixed effect estimate for the test position obtained from the simulations ($M_{sample} = -0.008$, HPD Interval: -0.017, -0.0002) converged with the estimate obtained from the model ($M_{model} = -0.008$). Moreover, MCMC p-value was 0.054 which suggests that the fixed test position effect approached significance. In summary, at the group level, the test position had a small but reliable effect on the false alarms, while the strength condition did not. The individual differences could be viewed in Figure 12, in which data from each individual was plotted along with the predicted false alarm rates, represented as lines.

REM As presented previously, REM does not predict the SBME after studying a list of items with mixed strength when the decision is solely based on the memory evidence. However, the question was whether incorporating the mechanism of encoding-at-retrieval would predict the SBME after a mixed study list. The model predictions obtained from REM is plotted in Figure 11. For simplicity, the parameters used in REM to simulate the hit rates and the false alarm rates in this mixed-study design were exactly the same as the parameters used in Experiment 1 (see Table 5). The predictions from REM showed an effect of the strength condition on the hit rates but not on the false alarm rates. Therefore, the SBME was not predicted by REM with the parameters used in this project. This prediction was consistent with the empirical findings of this experiment. However, it might be possible to get the SBME after a mixed-study list in REM with a different parameter set. In the parameter settings used here, encoding was better at test than at study and the u value at test was even higher than the strong study condition, which effectively causes every test trial to behave as a "strong" study trial. However, consistent with the design and the empirical results from this experiment, REM did not predict the SBME but predicted OI both in the hit rates and the false alarm rates. Moreover, the model did not predict an interaction of strength and test position for neither the hit rates nor the false alarm rates, which was also consistent with the data observed from the experiment.

The unreliable interaction between strength and test position effects for hit rates could stem from the similar levels of false alarm rates across two strength conditions. According to REM, hit rates decrease as a function of test position due to the updating of an incorrect trace and the addition of new traces to memory. When a foil item receives an 'old' response, the best matching trace will be updated with the foil item. As a result, if the target corresponding to that updated memory trace is tested later in the test sequence, the match between the trace and the target item will be relatively low. Since in this experiment, false alarm rates are similar in proportions across test conditions (strong test vs. weak test), the effect of false alarm rates on OI observed in hit rates will be similar in

magnitude. When the false alarm rates are higher for the weak foils compared to the strong foils as in Experiment 1, the decrease in the hit rates across test positions are expected to be (and in fact were) more prominent for the weak lists.

To summarize, the accuracy analysis from Experiment 2 shows that when participants study a list of items with mixed strength and are later tested on only the weak or the strong targets, they do not show the SBME. From the parameters used in this project, REM did not predict a SBME due to differentiation at retrieval when the mechanism of encoding at retrieval was implemented. Additionally, since participants were not informed on the strength of the test list prior to the test, the criterion did not shift. The root cause of a criterion shift is not well specified; some propose the criterion is set on the basis of the study items (e.g., Hirshman, 1995) and others propose that the criterion is set on the basis of the first few test items (e.g., Benjamin & Bawa, 2004). The pattern of data observed here are not consistent with a model based on estimating memory from the first few test trials. Starns et al claim that the SBME in both mixed- and pure-study lists is due to a change in the criterion. However, we do not find a SBME for the mixed-study list condition, suggesting that the simple Starns et al explanation is invalid. In Experiment 3, we attempt to replicate the Starns et al data by informing participants about the test conditions. But first, we evaluate OI. The item-noise models were supported by the OI observed in hit rates and false alarm rates. In the next section, the item-noise models and the attention hypothesis of OI will be further tested by using the DM.

Reaction Time and Diffusion Model Analysis

In this analysis, the effects of test list strength (strong and weak), the test position (5 test blocks) and the item type (target and foil) on reaction times of correct and responses were analyzed by applying the DM.

Reaction Time The same exclusion criteria were used as in Experiment 1. Figure 13 plots the reaction time quantiles as a function of response proportions for each condition. The vertical points are the 0.10, 0.30, 0.50, 0.70 and 0.90 quantiles of the reaction times averaged over participants. The darkest points represent the first test position bin (the average test position is 15) and as the shade gets lighter, the test position increases.

The test list strength effect can be observed in the hit rates both for accuracy and the reaction time distributions: Both in the spread of the reaction time quantiles (y-axis) and the change in response proportion (x-axis). The OI can also be observed for targets as a decrease in the spread of the reaction time quantiles, as well as an increase in the accuracy in the hit rates as test position gets shorter. In order to explain the processes that produce these data, we fit the DM to both group and individual data.

Diffusion Model Analysis The DM was fit to the reaction time and accuracy data by using the same methods as in Experiment 1. The first model (DM #1) tested the SBME by varying the drift rate and the starting point parameters across strength conditions. The item-noise account of OI was tested by varying the drift rate parameters across test positions. This model is exactly the same as the DM #1 fit in Experiment 1. There were a total of 27 free model parameters (20 from v, 2 from z, a, T_{er}, s_T, s_z and η) and the degrees of freedom in the data was 220.

The second model (DM #2) tested whether the decrease in the attention towards the end of the test list also contributes to OI in addition to the decrease in memory evidence. This was achieved by relaxing the constraint on the boundary separation parameter across test blocks, since the attention hypothesis predicts faster and more inaccurate responses. Similar to DM #2 in Experiment 1, the response bias was fixed across test blocks and specified to be different across strength conditions. The specific bias value was obtained from DM #1 and fed into DM #2. There were 29 free parameters in the second model (20 from v, 5 from a, T_{er}, s_T, s_z, η, 0 from z because z was estimated

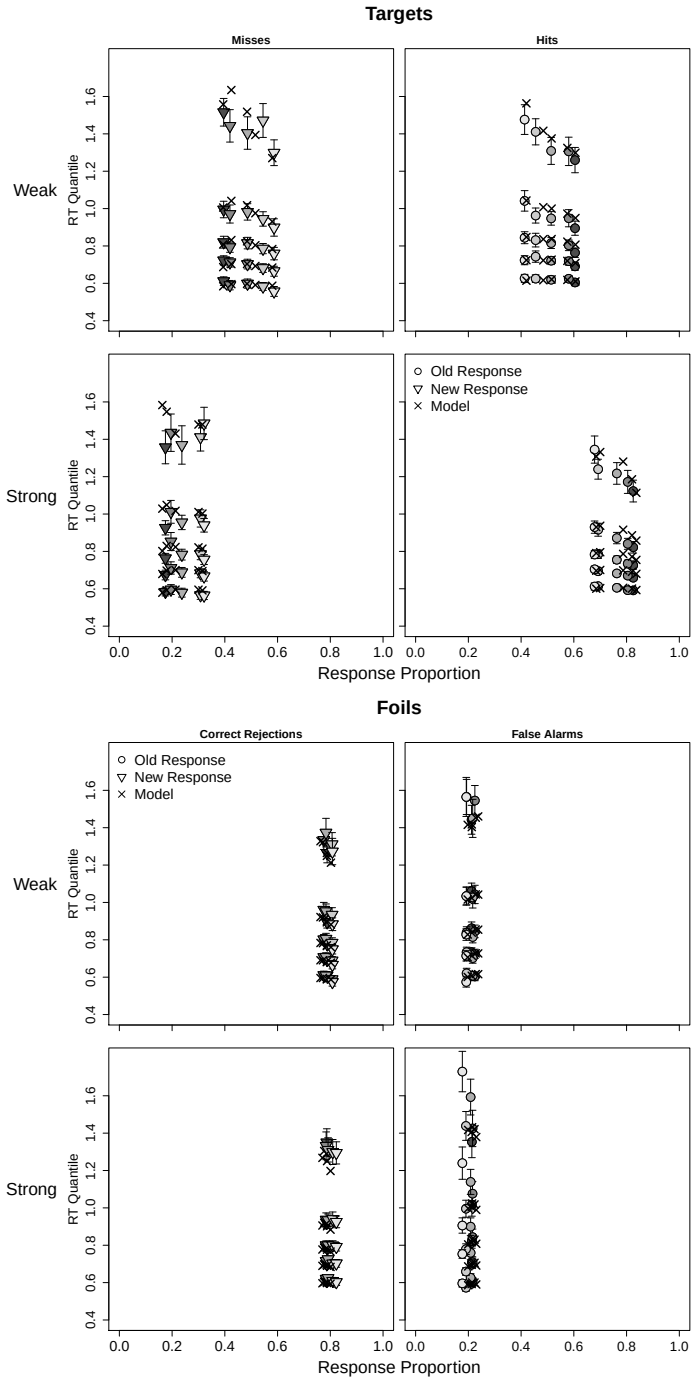

Figure 13: The reaction time distributions for Experiment 2 presented as the reaction time quantiles as a function of accuracy. The vertical plots corresponds to each condition which is a combination of the test block, the strength condition, the item type and the response. The test block is represented by the shades of grey: The darkest point is the first test block (the average test position is 15), one shade lighter grey point represent the second test block (the average test position is 45) and so on. Predictions from DM #2 (group data) is represented as $\times$.

Model	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)	χ^2 (Ind.)
1	1102	27	210941	0.00	211176	0.00	314
2	1061	29	210909	1.00	211162	1.00	304

Table 12: The diffusion model comparisons in Experiment 2. The χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion with finite sample correction, w_i(AICc) is the Akaike weights which represents the probability of the model being the best model, BIC is the Bayesian Information Criterion, w_i(BIC) is the Schwarz weights which represents the probability of the model being the best model according to the BIC values. The last column is the average χ^2 value from the fits to the individual data.

relative to *a*) [4]. Note that the differentiation account does not predict a difference in drift rate for foils between the strength conditions. However, both models permit the drift rate to vary in order to allow a direct comparison across the experiments.

Model fit statistics are presented in Table 12. AICc and BIC favored DM #2, which suggests that both a decrease in memory evidence and attention play a role in OI. The individual model fits to data from each participant also showed that DM #2 was more preferable for thirteen participants, DM #1 was more preferable for six participants and the model selection was inconclusive for seven participants. The parameters from the group data fits of the two models are presented in Table 13. The averaged parameters across the individual model fits are presented in Table A.14.

In order to test the SBME after a study list with mixed strength, the drift rate parameter was allowed to vary across strength conditions for targets and foils. A 2 (Strength) $\times$ 5 (Test Block) repeated measures ANOVA was conducted to analyze the drift rate estimates obtained from DM#2 that was fit to individual data. ANOVA results revealed a significant effect of strength on target drift rates, $F(1, 240) = 64.52$, $p < .001$. The mean drift rate for the strong targets ($M = 0.23$, $SD = 0.17$) was greater than the mean drift rate for the weak targets ($M = 0.038$, $SD = 0.13$). These parameter values show that the rate of evidence accumulation for strong targets was faster than that of weak

[4]Additional models were also fit to the group data to test the effects of the drift criterion and the starting point parameters separately. These models are presented in Appendix B along with a comparison across experiments.

Group Parameter Values (DM #1)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.121	0.357	−0.249	-0.245	0.178	
2	0.093	0.315	−0.213	-0.242	0.178	
3	0.037	0.262	−0.223	-0.249	0.178	
4	0.004	0.178	−0.246	-0.263	0.178	
5	−0.065	0.167	−0.272	-0.290	0.178	
Fixed Parameters	z/a_W 0.47	z/a_S 0.49	η 0.209	s_z 0.156	T_{er} 0.580	s_T 0.294
Group Parameter Values (DM #2)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.120	0.356	−0.247	-0.244	0.176	
2	0.093	0.318	−0.215	-0.246	0.182	
3	0.038	0.265	−0.224	-0.257	0.182	
4	0.004	0.178	−0.246	-0.264	0.178	
5	−0.063	0.168	−0.270	-0.287	0.175	
Fixed Parameters	z/a_W 0.47	z/a_S 0.49	η 0.207	s_z 0.157	T_{er} 0.580	s_T 0.295

Table 13: The group parameter values of DM from Experiment 2. In the first model, the boundary separation was fixed across the test blocks.

targets, consistent with the results from the accuracy analysis and REM. The critical finding from the accuracy analysis was the null effect of the test strength condition on the false alarms rates in this experiment. The ANOVA results did not reveal a significant effect of strength on the drift rates of foils. Figure 14 plots the box-and-whisker diagrams of the drift rates as a function of item type, test list strength and test position [5]. Evidence for the item-noise models account for OI comes from the decrease in the drift rate parameter values for the targets as a function of test block, $F(4, 240) = 5.40$, $p < .001$. The mean drift rate at the first test block ($M = 0.22$, $SD = 0.19$) was significantly greater than the

[5]The drift rates for target and foils were compared across experiments and the results of these comparisons are presented in Appendix A

mean drift rate at the last test block ($M = 0.04$, $SD = 0.15$), $t(51) = 8.79$, $p < .001$. The effect of test block on the drift rate of foils was not significant neither was the interaction between the strength condition and test block.

The response bias (z/a) estimated in DM #1 showed that the participants were more inclined overall to respond 'no' as the values were lower than 0.5. In this experiment, participants were not told about the strength of the list that they would be tested on thus the starting point parameter was not expected to differ across strong and weak test lists. The pairwise t-test results did not show a significant effect of strength on the relative starting points obtained from each participant (Figure 15).

Finally, the boundary separation parameter was estimated to change as a function of the test block in DM #2 according to the attention hypothesis of OI. In order to test this hypothesis a one-way repeated measures ANOVA was conducted on the boundary separation parameter estimates obtained from DM#2 that was fit to the individual data. The results from ANOVA did not show a significant effect of test block on the boundary separation parameter. Thus, the boundary separation parameter values estimated in this experiment did not provide evidence for the attention hypothesis to account for OI. Although, the ANOVA results from the parameter estimates of the individual fits of DM#2 showed a null effect of test block, the model fitting statistics from the group data preferred DM#2. This discrepancy could be due to the overfitting of noise in the group data.

To summarize, as the item-noise models would predict, the drift rate parameter values for the strong targets were greater than those of the weak targets. The strength effect on the drift rates for foils was not consistent with the results from the accuracy data and predictions from REM. Finally, the decrease in the drift rates as a function of the test position provided evidence for the item-noise models, as they predict an increase in interference towards the end of the test list. However, the attention hypothesis account for the explanation of OI was not supported, because the boundary separation parameter did not show a consistent pattern across test position.

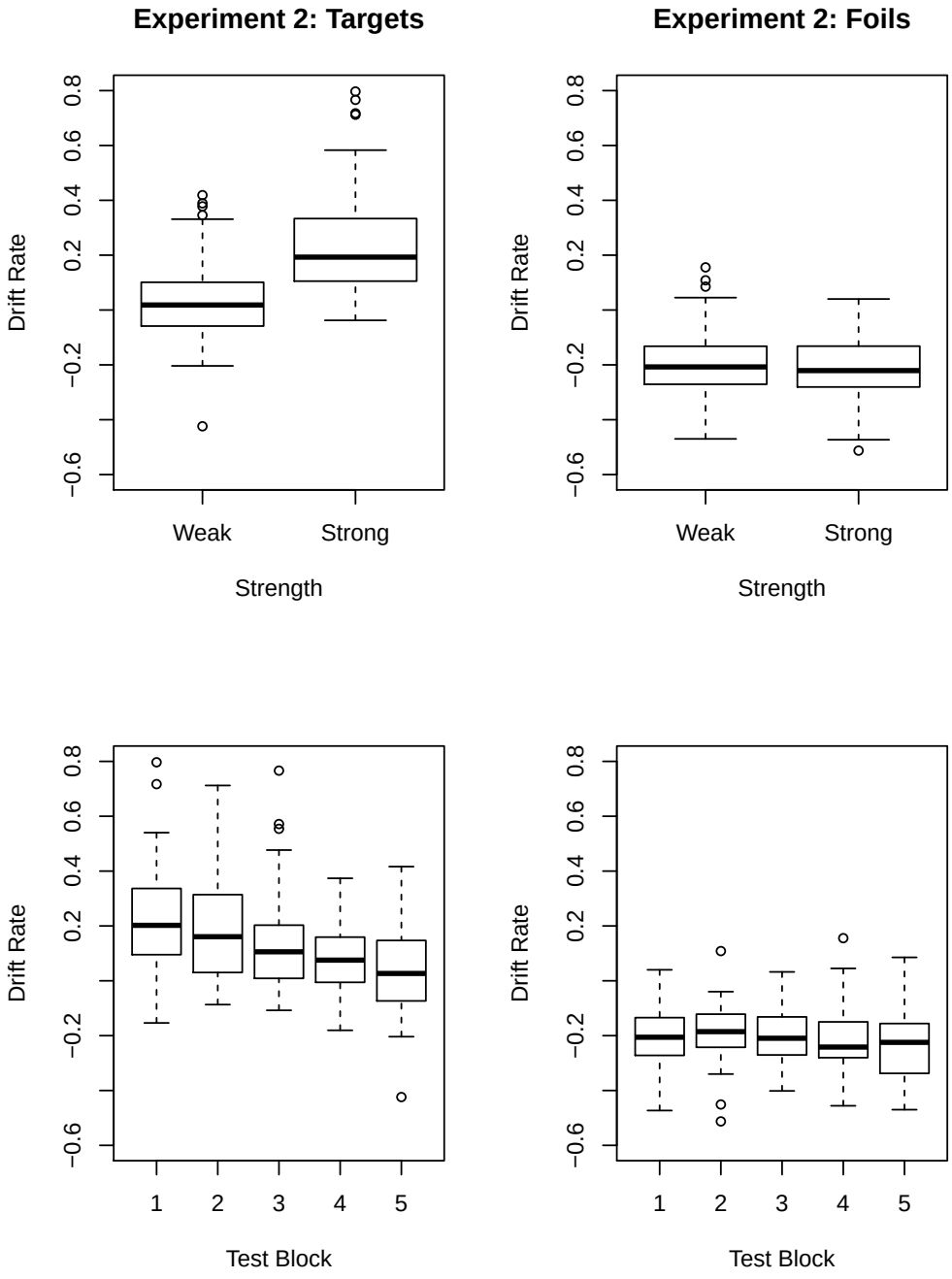

Figure 14: Box-plots for the drift rate parameter estimates obtained from DM#2 that was fit to individual data. The first row plots the drift rate estimates from each individual as a function of the strength condition for targets and foils. The second row plots the drift rate estimates as a function test block.

Summary

The motive of Experiment 2 was to test whether the SBME observed after studying a mixed strength list is due to differentiation at retrieval. The results of the experiment

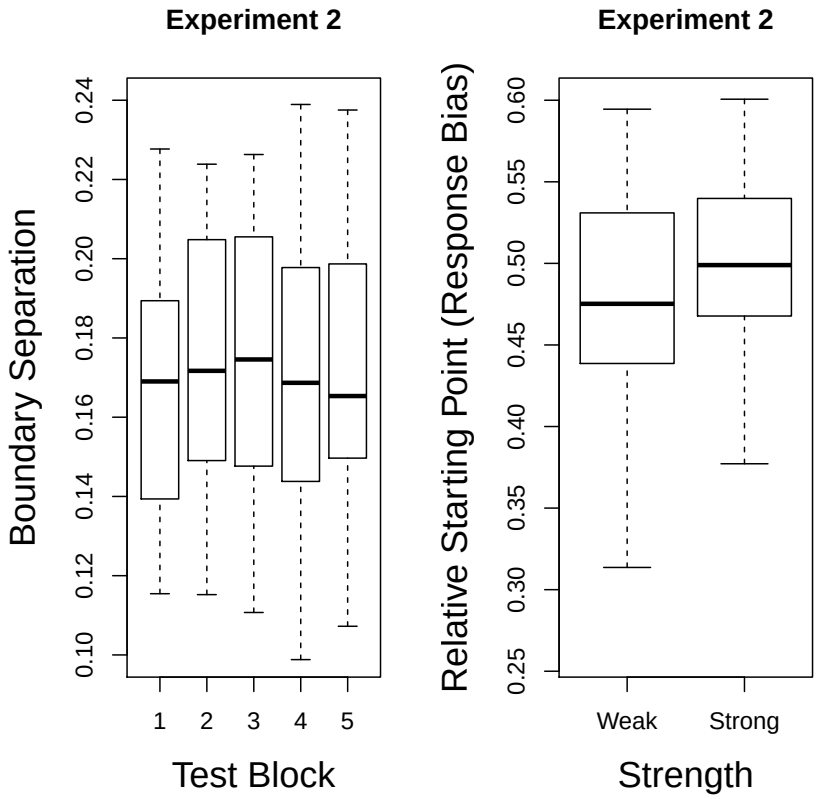

Figure 15: On the left, box-plots for the boundary separation parameter estimates obtained from DM#2 that was fit to individual data. On the right, box-plots for the response bias estimate from DM#1 that was fit to each individual.

showed that when participants were not informed on the strength of the targets that they would be tested on, the SBME was not observed. Moreover, REM accurately predicted the data using the same parameters as Experiment 1. Thus, these findings suggest that the SBME observed after a mixed strength study list is more likely due to a change in the criterion placement because participants are informed on the strength of the test lists. This hypothesis will be tested in the next experiment. As for the findings for OI, a decrease in the hit rates and the drift rates for targets as a function of test position provided evidence for the item-noise explanation of OI.

Experiment 3

In order to further test the criterion shift account, the Starns et al. (2012) approach of informing participants of the target strength was replicated in this experiment. Participants studied a list of items with mixed memory strength and each test list was constructed from either the strong or weak targets along with the new items. Different from the second experiment, participants were *informed* of the encoding task that the targets were subjected to during study. Thus, in this experiment participants were given the opportunity to set a different criterion to endorse items across test list conditions. The criterion shift account predicts the SBME such that hit rates would be higher, and false alarm rates would be lower for the strongly encoded items because when participants are tested on the strong list, they would require more evidence to endorse an item. That same criterion shift account implemented in the REM framework makes the same prediction. However, without a change in criterion, REM does not predict a difference in the false alarm rates across test list strength conditions.

Methods

Participants

Thirty-one Syracuse University undergraduates took part in the experiment in exchange for course credit. Fifteen of the participants were female. Participants who had overall low accuracy level ($d' < 0.5$) were excluded from the subsequent analysis. The resulting sample

size was twenty-two participants.

Materials

The words were sampled from the same word pool used in Experiments 1 and 2.

Procedures and Design

The procedures were exactly the same as in Experiment 2 except that the participants were informed at the beginning of the test on which type of targets they will be tested on. At the beginning of weak test lists, the information about the targets were displayed as "You will be tested only on the words for which you decided whether they contain the letter 'e' ." For the strong test lists, the information on the screen was "You will be tested only on the words for which you made a pleasantness judgment." The information was on the screen for 5 sec in both types of test lists. At the beginning of the experiment, participants were instructed to pay attention to the information that they would receive before the test trials begin. The experiment was a 2 (Strength) $\times$ 5 (Test Block) within subjects-design. Test positions were binned into 5 test blocks and when the responses from all of the blocks were pooled with respect to the strength condition, each test block contained 180 trials for each participant.

Results and Discussions

Accuracy

The effects of the strength condition and the test position on the hit rates and the false alarm rates were assessed by a linear mixed effects model analysis. Initially, the intraclass correlation was calculated to confirm the need to apply the mixed effects model both for the hit rates and the false alarm rates. Later, models with different complexities were fit to

the hit rates and the false alarm rates separately. The model that explains best the variance in the data with minimum complexity has been selected as the final model. Hit rates and false alarm rates were simulated by using REM to provide an explanation for the accuracy results obtained from this experiment.

Hit Rates The intraclass correlation obtained from the intercept-only model was 44% which shows a reliable variability between participants. In the subsequent models, first strength was added as a fixed effect (Model 2) and, later also as a random effect (Model 3). A decrease in AIC and BIC, and the χ^2 difference test suggested that adding the strength condition as a predictor which is also allowed to differ for individuals improved the model by deceasing the unexplained variance. Later, the test position was added to the intercept-only model as a fixed effect (Model 2a) and also as a random effect (Model 3a) to assess the OI in hit rates. The model comparisons suggested that only the fixed effect of the test position is reliable. However, when the test position effect was later added as a fixed and a random effect to the strength model (Model 4 and 5), both the test position and the strength condition effects were reliable at the group level. In addition to that, these effects were required to differ for each individual due to the between-subject variability. Finally, the interaction term was tested by adding it as a fixed (Model 6) and a random effect (Model 7). When the interaction term was added as fixed effect, the model fit did not improve significantly. On the other hand, when added as a random effect the unexplained variance in the hit rates decreased significantly according to AIC, BIC and the χ^2 difference test. However, in the full model (Model 7), the fixed parameter estimate for the interaction was not reliably different than 0 (0.009, $SE = 0.006$,$t = 1.35$). As a result, a reduced model (Model 8) was fit to the data and the model comparison statistics showed that the model was as powerful as Model 7 in explaining the variance (see Table 14 for model comparisons).

Parameter estimates from the best fitting model (Model 8) are presented in

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-125.86	-115.67	3	65.928	
2	Intercept, Strength	Intercept	-264.71	-251.14	4	136.357	M2-M1=140.86*
2a	Intercept, Test Position	Intercept	-154.37	-140.79	4	81.184	M2a-M1=30.513*
3	Intercept, Strength	Intercept, Strength	-316.88	-296.52	6	164.44	M3-M2=56.169*
3a	Intercept, Test Position	Intercept, Test Position	-150.37	-130.01	6	81.187	M3a-M2a=0.006
4	Intercept, Strength, Test Position	Intercept, Strength	-446.72	-422.96	7	230.36	M4-M3=131.84*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-457.46	-423.53	10	238.73	M5-M4=16.744*
6	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position	-458.60	-421.27	11	240.30	M6-M5=3.13
7	Intercept, Strength, Test Position, Interaction	Intercept, Strength, Test Position, Interaction	-465.03	-414.13	15	247.51	M7-M6=14.428*
8	Intercept, Strength, Test Position	Intercept, Strength, Test Position, Interaction	-465.21	-417.70	14	247.51	M7-M8=1.817

Table 14: Model comparisons for hit rates (Experiment 3). Estimates are based on 220 data points. AIC is Akaike Information Criterion, BIC is Bayesian Information Criterion, log-lik is log-Likelihood. The significant differences between the χ^2 values of each model pair at $\alpha = 0.05$ level is represented by * sign.

Table 15. The fixed effects show that participants were more likely to endorse the strongly encoded targets compared to the weakly encoded targets. The slope of the test position effect provides evidence for the output interference observed in the hit rates ($b = -0.04$,

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.60	0.03	22.30	0.17
Strength	0.23	0.03	7.49	0.18
Test Position	-0.04	0.004	-8.97	0.025
Strength × Test Position	-	-	-	0.02
Residual	-	-	-	0.05

Table 15: The best fitting mixed model parameters for the hit rates (Model 8) in Experiment 3. The strength effect has been treated as a factorial variable so the fixed effect estimate of the intercept is the estimated hit rate for the weak targets when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

$SE = 0.004$, $t = -8.97$). Hit rates are plotted as a function of the strength condition and the test position in Figure 16. The points represent the hit rates averaged over participants and the solid lines represent the predicted values from Model 8 for the hit rates. The random effects for the interaction term in Model 8 suggest that the test position effect depends on the strength of the test list differently across participants. Figure 17 plots data separately for each individual along with the best fitting model predictions. For example, while some participants showed a more prominent decrease in the hit rates of the weak targets (e.g., participant 21), others showed a more prominent decrease in the hit rates of the strong targets (e.g., participant 16). It could also be seen from the figure that most of the participants did not show an interaction as the rate of decrease in the hit rates was similar across the strength conditions (e.g., participants 3, 5, 7, 9, 10 and so on). Thus, when the individual effects of the interaction term are averaged over participants, the results do not provide a consistent effect of the interaction term at the group level.

False Alarm Rates The intercept-only model fit to the false alarm rates produced an intraclass correlation of 71% which shows that a very large amount of variance in the false alarm rates can be explained by the individual differences. To further explain the variance, the effect of the strength condition was added as a fixed effect (Model 2) and also a

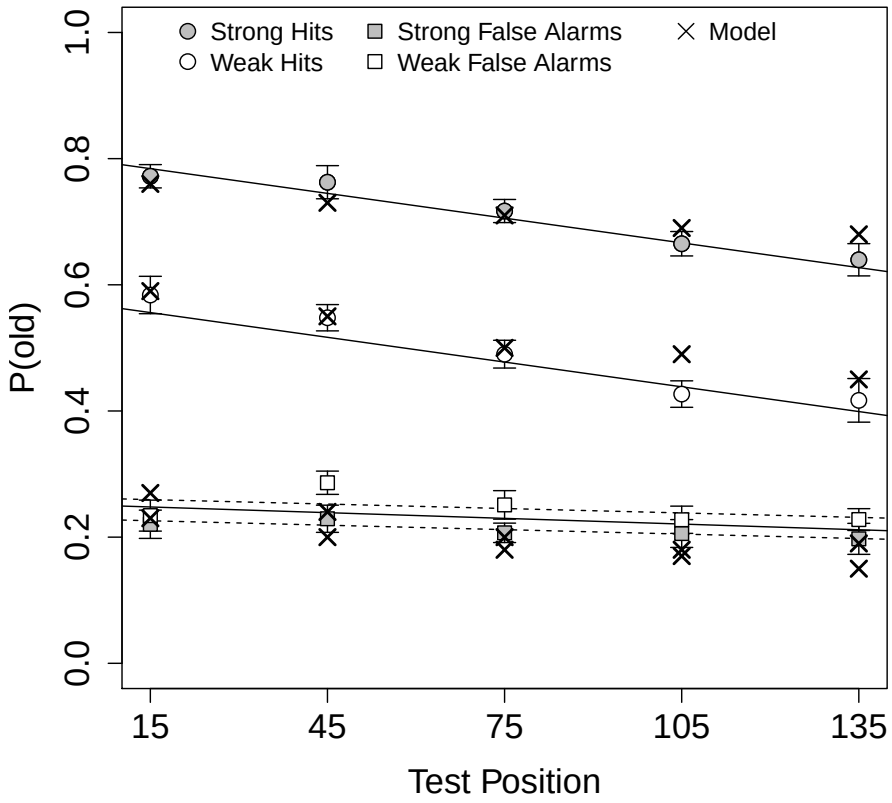

Figure 16: The hit rates and the false alarm rates as a function of the strength condition and the test position in Experiment 3. The lines represent the predictions from the best fitting linear mixed effects model (Model 8 for the hit rates and Model 6 for the false alarm rates). The dashed line represents the predicted false alarm rates from an alternative model for the false alarm rates (Model 5). The squares and the points represent the false alarm rates and the hit rates averaged over participants. Error bars are the within-subjects 95% CI. $\times$ represent the predicted hit rates and the predicted false alarm rates from REM.

random effect (Model 3). The results from the model comparison statistics demonstrate reliable fixed and random effects of the strength condition on the false alarm rates (Table 16). The model fits improved when the test position was also added as a predictor (Model 2a) but not with a random effect (Model 3a). To assess the combined effects of the test position and the strength condition, both predictors were added as fixed and random effects (Models 4 and 5).

In Model 5, the fixed effect parameter of strength is -0.033(SE=0.018, t=-1.89). As

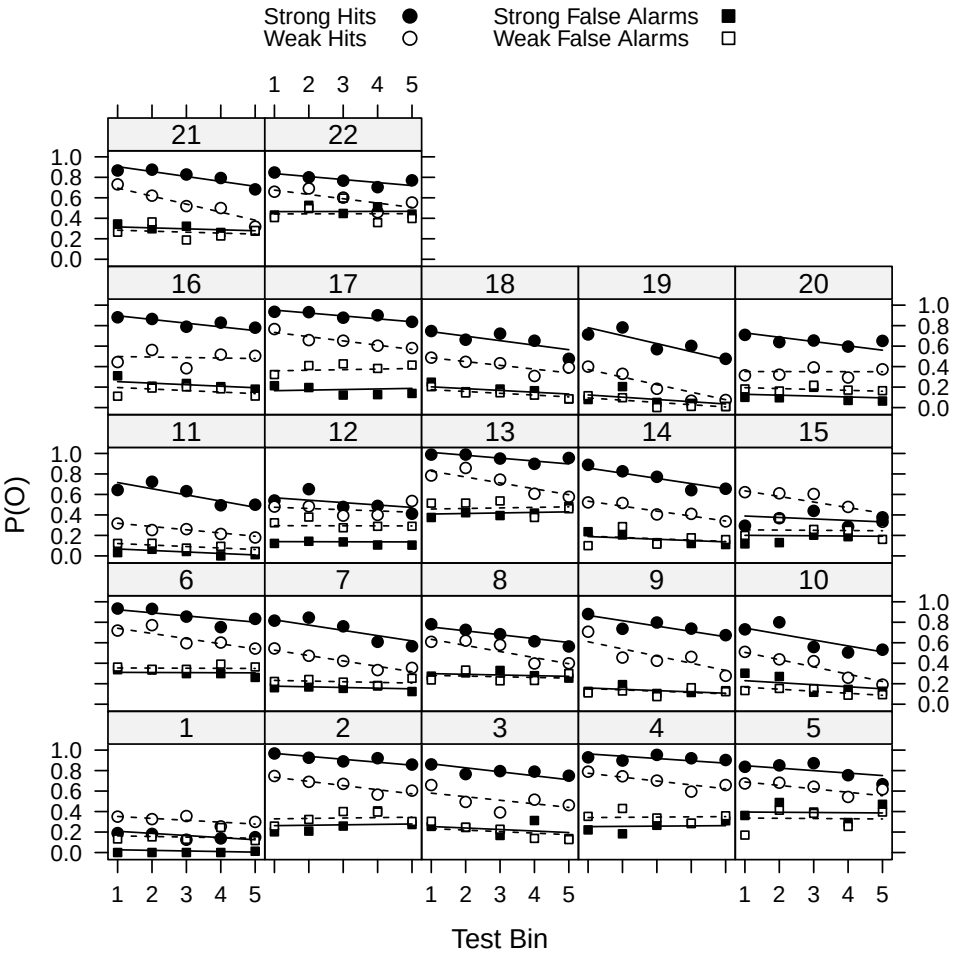

Figure 17: The hit rates and the false alarm rates as a function of the strength condition and the test position for each individual participant (Experiment 3). Data are represented by the points and the squares. The linear mixed effect model predictions are represented by the lines (Model 8 for the hit rates and Model 6 for the false alarm rates).

a result, to test the reliability of the strength effect, a reduced model was compared to Model 5. In the reduced model (Model 6), strength was added only as a random effect and removed from the fixed effect predictors. The comparison of the reduced model and the full model is presented in Table 16. The χ^2 difference test suggested that removing the strength effect did not harm the model according to the alpha level of 0.05 as the p-value for the difference was 0.067. However, this p-value suggests that the χ^2 difference between the models approaches significance. Consistent with that, the t-value of the fixed effect for the strength condition from Model 5 suggests a trend for the strength effect in false alarm

Model	Fixed Effects	Random Effects	AIC	BIC	df	log-Lik	χ^2 Difference Test
1	Intercept	Intercept	-468.47	-458.29	3	237.24	
2	Intercept, Strength	Intercept	-479.43	-465.86	4	243.72	M2-M1=12.96*
2a	Intercept, Test Position	Intercept	-470.96	-457.38	4	239.48	M2a-M1=4.485*
3	Intercept, Strength	Intercept, Strength	-515.71	-495.35	6	263.85	M3-M2=40.276*
3a	Intercept, Test Position	Intercept, Test Position	-470.23	-449.87	6	241.12	M3a-M2a=3.275
4	Intercept, Strength, Test Position	Intercept, Strength	-520.83	-497.08	7	267.42	M4-M3a=7.122*
5	Intercept, Strength, Test Position	Intercept, Strength, Test Position	-524.20	-490.26	10	272.10	M5-M4=9.365*
6	Intercept, Test Position	Intercept, Test Position, Strength	-522.75	-492.20	9	270.37	M5-M3c=3.448

Table 16: Model comparisons for the false rates (Experiment 3). Estimates are based on 220 data points. AIC is Akaike Information Criterion, BIC is Bayesian Information Criterion, log-lik is log-Likelihood. The significant differences between the χ^2 value of each model pair at $\alpha = 0.05$ level is represented by * sign.

rates. In order to test the reliability of the fixed effect of the strength condition, a Monte Carlo Chain Model simulation was used for estimating the parameters of Model 5. The estimates are based on 10000 MCMC samples generated from the posterior distribution of the model parameters. The mean fixed strength effect obtained from the simulations converged with the estimate obtained from Model 5 ($M_{sample} = M_{model} = -0.034$, HPD Interval for sample estimate: -0.059, -0.007, MCMC p-value = 0.013). These results suggest a tendency for a strength effect in false alarm rates as the false alarm rate for the strong list was lower than the false alarm for the weak list. Finally, the false alarm rates were

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.27	0.02	12.00	0.09
Test Position	-0.007	0.003	-2.151	0.01
Strength	-0.34	0.02	-1.89	0.08
Residual	-	-	-	0.05

Table 17: An alternative mixed effects model parameters for false rates in Experiment 3 (Model 5). The intercept is the estimated hit rate when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

found to decrease as a function of test position as a slight effect of the test position was observed in the false alarm rates.

In Figure 16, the predicted false alarms rates from both of the models were plotted as a function of the strength condition and the test position. The solid line represents the predicted values from Model 6 and the dashed lines represent the predicted values from Model 5. Additionally, the individual variances on the strength effect can be observed from Figure 17. The figure shows that while some participants adapt to a different criterion for the strong and the weak test lists (e.g. participants 1, 12 and 17) and show the SBME, others do not show a strength effect on the false alarm rates (e.g. participants 8, 9, 14 and 22). These results suggest that when participants were informed of the strength of the targets that they will be tested on, some adopt a different criterion and some do not. This is also evident from the inconclusive model selection such that the data did not discriminate between Model 5 and 6. Parameter estimates of Model 5 and 6 are presented in Table 17 and Table 18 respectively.

REM As presented in the second experiment, REM does not predict the SBME on the basis of differentiation after the encoding-at-retrieval mechanism was implemented with the parameters that would account best for the OI observed in the experiment. In this experiment, the criterion shift explanation of the Starns et al. (2012) paradigm was tested.

Predictor	Fixed Effect Estimate	Fixed Effect Standard Error	t value	Random Effect Estimate (Standard Deviation)
Intercept	0.23	0.02	10.07	0.12
Test Position	-0.008	0.004	-2.42	0.01
Strength	-	-	-	0.05
Residual	-	-	-	0.05

Table 18: Best fitting mixed model parameters for false rates in Experiment 3 (Model 6). The intercept is the estimated hit rate when the test position is equal to zero. t-values less than -2 or greater than 2 suggest that the fixed effect is reliably different than 0.

Thus, in order to account for the tendency for a strength effect in false alarm rates, the criterion shift account was applied in REM. In addition to varying u parameter across item strength during study, *criterion* parameter was also varied across test list strength. All of the parameters used in this simulation were identical to the parameters used in the simulations of the previous experiments except for the *criterion* parameter (Table 5). The *criterion* parameter was set to 0.70 for the weak test list and 0.80 for the strong test list as the criterion shift account posits that participants require more evidence to endorse an item when they are presented with the strong targets in the test list. Therefore, they set a stricter criterion to provide an 'old' response compared to the criterion that they set at the beginning of the weak test lists. The predicted hit rates and the predicted false alarm rates are presented in Figure 16. The model captured the strength effect observed in the hit rates as the hit rates of the weak targets were lower than the hit rates of the strong targets. Similarly, OI was also predicted for the hit rates which decreased as a function of the test position. The slight strength effect on the false alarm rates was manifested compared to the predicted false alarm rates from Experiment 2. The slight OI effect in false alarm rates was also predicted by REM. Thus, the results from REM simulations show that a change in the criterion for willingness to endorse a test item can predict the SBME in a mixed-study list paradigm.

To summarize, the results from the accuracy data of Experiment 3 show that when

the participants were given the information about the strength of the targets that they will be tested on, they may set a different criterion with more evidence required to endorse items from a well encoded list. In this experiment, strength was manipulated by a levels-of-processing task in which the strong items were encoded by a deep processing and the weak items were encoded by a shallow processing task. The participants were informed about the strength of the test list by being provided with a description of the task they used for encoding the item during the study (e.g. the pleasantness or the letter task). Thus, the information that participants received at the beginning of the test was perhaps not as straight forward as being presented by the number of times the items were studied as in the Starns et al. (2012) paradigm. For example, participants are aware from a lifetime of experience that additional encoding time helps memory, however they are less likely to be aware that a judgment of pleasantness in comparison to judging the letter 'e' lead to different levels of accuracy. However, some participants showed a strength effect in the false alarm rates. On average, there was a tendency for a strength effect on the false alarm rates but this effect was not as prominent and as reliable as the strength effect in Experiment 1. Thus, these results suggest that when the criterion shift is the only source of the SBME, the strength effect on the false alarm rates is not as strong as it would be after studying lists with pure strength. Additionally, the strength effect in this experiment was also found to be highly variable across participants. One possible reason for the high variability in the strength effect is that many of the participants were taking their first psychology course and while most were unlikely to be aware of the depth of encoding effect on memory, others may have learned about the effect in class. In the following section, the reaction time data was analyzed by the DM to further test the item-noise models, the criterion shift account and the attention hypothesis for OI.

Reaction Time and Diffusion Model Analysis

In this section, the effect of the test list strength (strong and weak), the test position (5 blocks), and the item type (target and foil) on the reaction time of correct and error responses were analyzed.

Reaction Time The reaction time analysis was identical to that in Experiment 2. The reaction time quantiles for this experiment are plotted in Figure 18. The figure shows that the shape of the reaction time changes across the strength and the test position conditions. Likewise, the response proportions also show a pattern consistent with the SBME and OI. In order to investigate how memory processes and decision processes contribute to these effects, the DM was fit to averaged group data and data from each individual.

Diffusion Model Analysis The DM was fit to the reaction time quantiles and accuracy data of 20 conditions (strength $\times$ test position $\times$ item type). In the first model (DM #1), the criterion shift account of the SBME was tested by varying the drift rate of targets and foils and the starting point parameter across the strength conditions. In a previous study, Starns et al. (2012) showed that the model that allowed both the drift rate and the starting point across the test list strength conditions provided the best fit to the data in comparison to the models in which only one of those parameters was free to vary[6]. In the current analysis, both of these parameters were allowed to vary and the change in the drift rate is assumed to reflect the shift in the drift criterion. In order to further test the item-noise account of OI, the drift rate parameter was allowed to vary across test positions since the item-noise accounts predict a decrease in the memory evidence as the number of tested items increase. There were 27 free parameters in this model (20 from v, 2 from z, a, T_{er}, s_T, s_z and η).

The second model (DM #2) tested the attention hypothesis of OI by varying the

[6]See the Appendix for the models in which either the starting point or the drift rate parameter was varied across the list strength conditions.

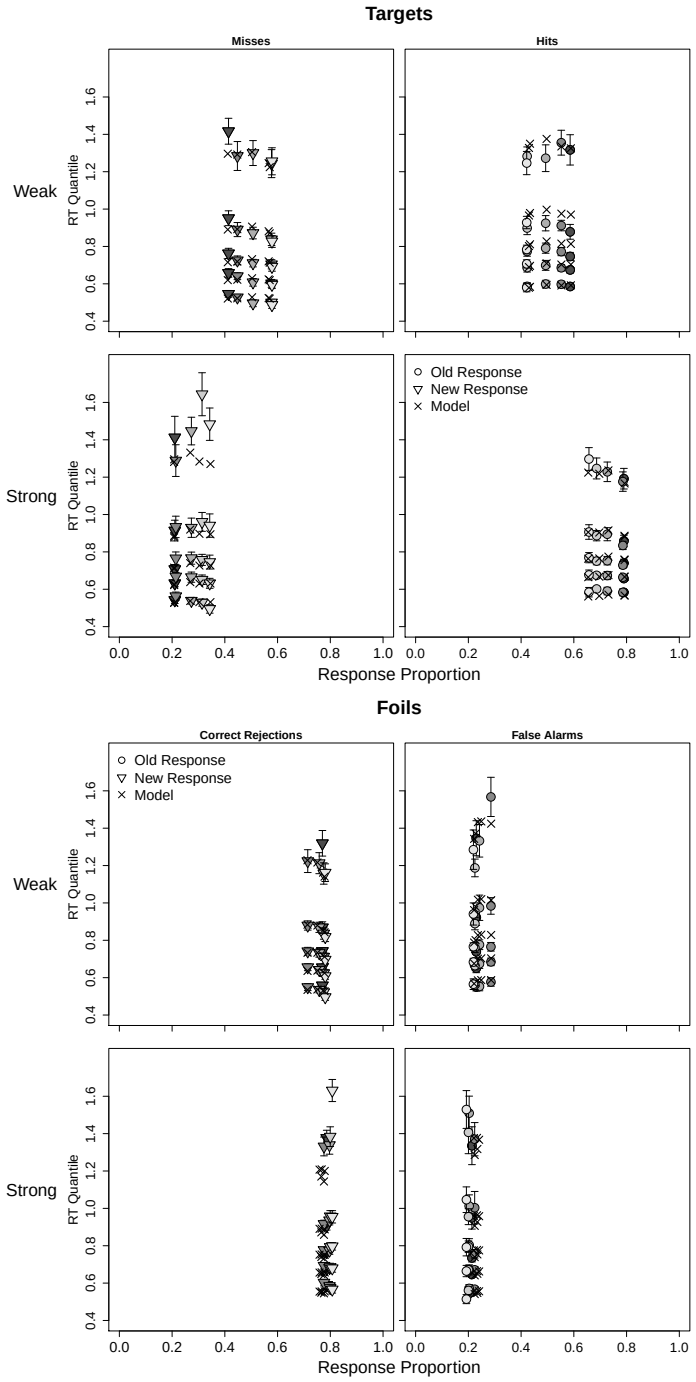

Figure 18: The reaction time distributions for Experiment 3 presented as quantiles as a function of accuracy. Each vertical plot represents one condition which is a combination of the strength condition, the test position and the item type. The test position condition is represented by the shades of grey as the darkest point is the first test block (average test position of 15) and the lightest point is the last test block (average test position of 135). Model predictions are from the DM #2.

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)	χ^2 (Ind.)
1	1440	27	178868	0.00	179099	0.00	350
2	1429	29	178806	1.00	179054	1.00	335

Table 19: The χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion with finite sample correction,w_i(AICc) is the Akaike weights which represents the probability of the model being the best model, BIC is the Bayesian Information Criterion, w_i(BIC) is the Schwarz weights which represents the probability of the model being the best model according to the BIC values. The last column is the average χ^2 value from the fits to the data from each participant.

boundary separation parameter across test blocks. As in the analysis of previous experiments, the response bias was constrained on the value (z/a) obtained from the first model fit. Thus, the changes observed in the boundary separation parameter was not confounded with a change in the response bias across test blocks. The number of free parameters was increased to 29 when the boundary separation was allowed to vary (20 from v, 5 from a, T_{er}, s_T, s_z and η).

Table 19 presents the model fit statistics for the two models described above. Both AICc and BIC favored DM #2. Fits of the models to individual participants showed that the data from ten participants favored DM #2, the data from eight participants preferred DM #1 and the model selection was inconclusive for four participants. The model parameter estimates from the group data fit are presented in Table 20 and the average parameter values from individual fits are presented in Table A.15.

A 2 (Strength) $\times$ 5 (Test Block) repeated measures ANOVA was conducted to test the effect of list strength and OI on the drift rate parameters from individual DM#2 fits. The results revealed a strength effect on the target drift rates, $F(1, 21) = 60.89$, $p < .001$. The mean drift rate of the strong targets ($M = 0.16$, $SD = 0.15$) was greater than the mean drift rate of the weak targets ($M = 0.05$, $SD = 0.10$). The drift criterion hypothesis predicts higher drift rates in absolute value for the strong foils. This prediction is due to setting a more stringent criterion for the accumulation of evidence when an item is tested along with the strong targets. Even when the memory evidence for a foil does not differ, if

Group Parameter Values (DM #1)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.052	0.236	−0.128	-0.172	0.152	
2	0.046	0.226	−0.096	-0.159	0.152	
3	0.004	0.198	−0.110	-0.189	0.152	
4	0.005	0.147	−0.147	-0.190	0.152	
5	0.009	0.126	−0.166	-0.193	0.152	
Fixed Parameters	z/a_W 0.43	z/a_S 0.48	η 0.244	s_z 0.129	T_{er} 0.548	s_T 0.195

Group Parameter Values (DM #2)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.089	0.200	−0.126	-0.155	0.148	
2	0.068	0.201	−0.092	-0.149	0.147	
3	0.034	0.151	−0.117	-0.163	0.148	
4	−0.001	0.132	−0.136	-0.158	0.143	
5	−0.003	0.106	−0.142	-0.168	0.141	
Fixed Parameters	z/a_W 0.43	z/a_S 0.48	η 0.228	s_z 0.120	T_{er} 0.539	s_T 0.153

Table 20: The group parameter values of DM from Experiment 3. In the first model, the boundary separation was fixed across the test blocks.

the foil is tested in a strong list, the memory evidence of that foil is less likely to pass that stringent threshold. Then, the memory evidence will accumulate faster towards the 'new' boundary. The strength effect on the foil drift rates approached significance from the ANOVA results, $F(1, 21) = 3.86$, $p = .06$. The mean drift rate of the strong foils was -0.17 ($SD = 0.07$) and the mean drift rate of the weak foils was -0.14 ($SD = 0.09$). Additionally, the test block had an effect on the drift rates of targets as the item-noise models would expect. ANOVA results showed that the target drift rates decreased as the test position increased, $F(4, 84) = 29.25$, $p < .001$. The mean drift rate of the targets at the first block ($M = 0.19$, $SD = 0.16$) was significantly greater than the targets at the last block

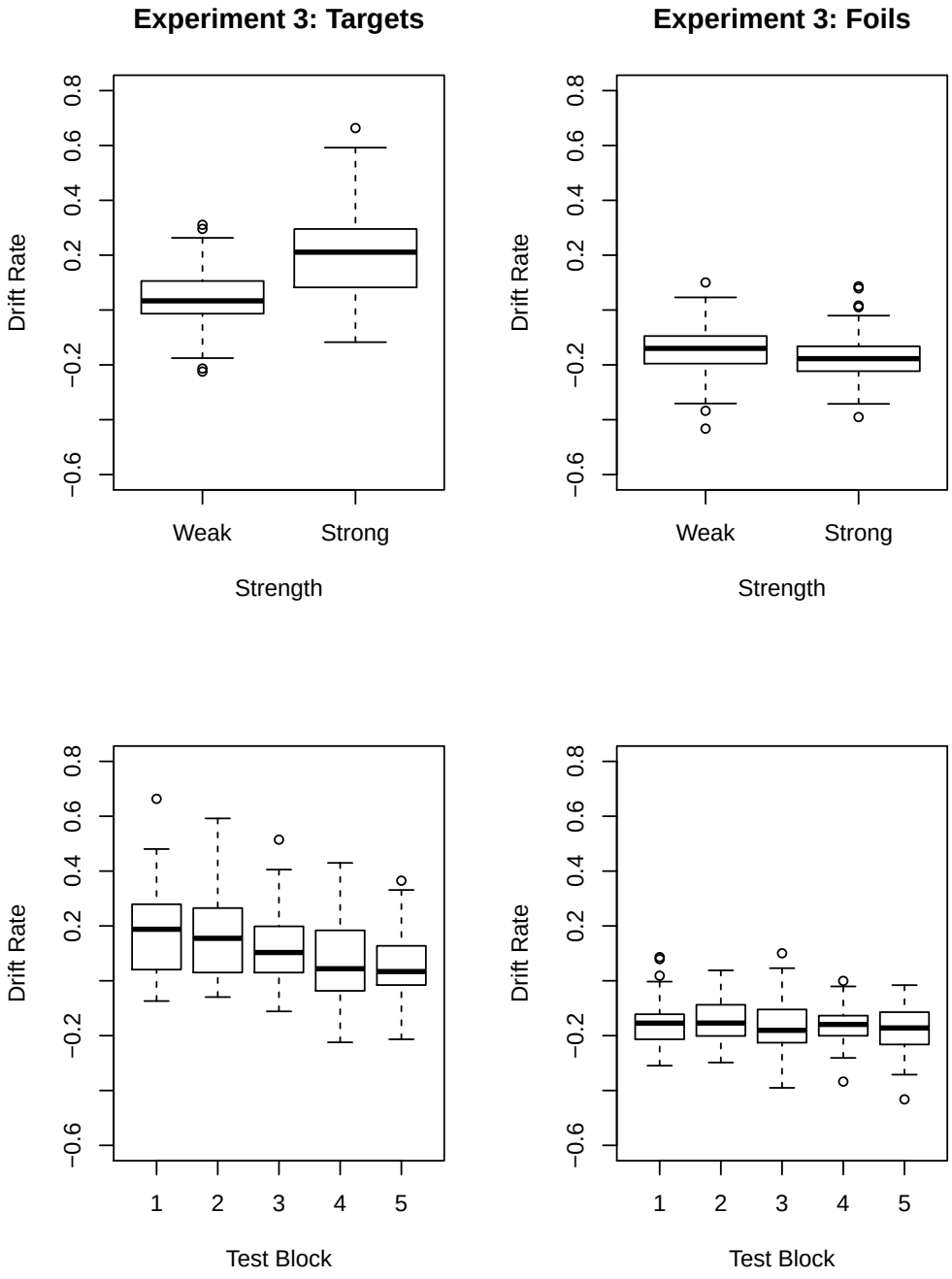

Figure 19: Box-plots for the drift rate parameter estimates obtained from DM#2 that was fit to individual data. The first row plots the drift rate estimates from each individual as a function of the strength condition for targets and foils. The second row plots the drift rate estimates as a function test block.

($M = 0.06$, $SD = 0.12$), $t(43) = 8.44$, $p < .001$.

To test the effect of strength on response bias, one-way repeated measures ANOVA

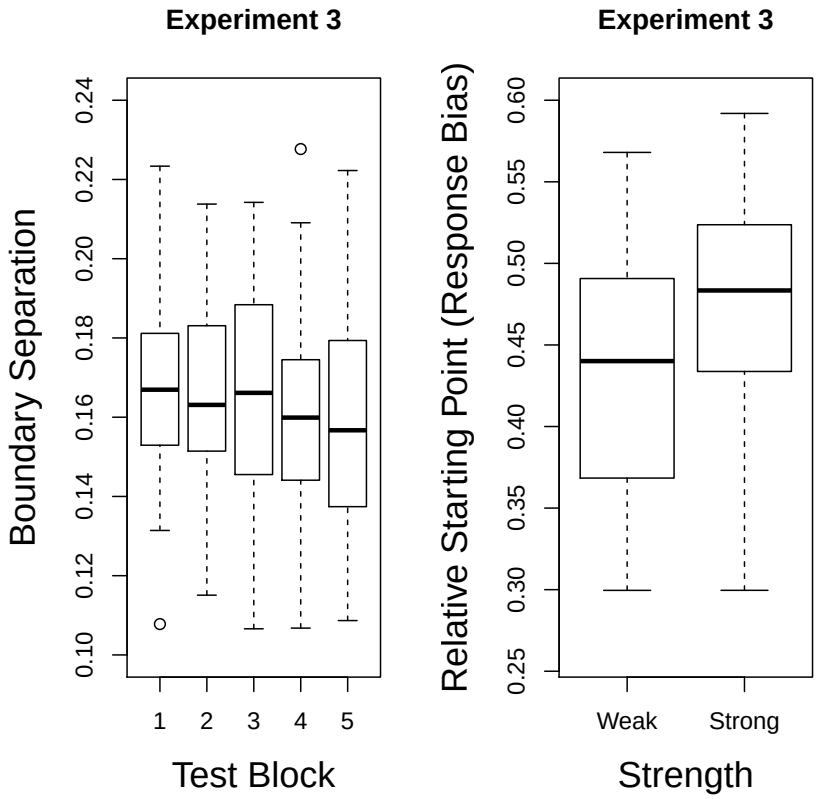

Figure 20: On the left, box-plots for the boundary separation parameter estimates obtained from DM#2 that was fit to individual data. On the right, box-plots for the response bias estimate from DM#1 that was fit to each individual.

was conducted on the response bias estimates obtained from DM#1 fits to individual data. The results revealed a significant effect of strength, $F(1, 21) = 9.19$, $p < .001$, as the participants were more likely to respond 'old' when they were tested with the strong lists ($M = 0.48$, $SD = 0.08$) in comparison to the weak lists ($M = 0.43$,$SD = 0.07$). These results suggests that participants were more likely to be biased to reject an item when the test list was constructed from the weak targets. In other words, informing participants the strength of the list that they will be tested on also affects the starting point parameter in the DM.

Finally, the boundary separation parameter value was found to be affected by the test block. One way ANOVA applied to individual boundary separation parameter

estimates from DM#2. The results revealed a significant effect of test block on the boundary separation parameter, $F(4, 84) = 5.46$, $p < .001$. Further contrasts showed that the effect was peculiar to the last two blocks. The boundary separation parameter estimate dropped significantly at the fourth block as the mean of the estimate values was significantly greater at the third block ($M = 0.166$, $SD = 0.03$) compared to the mean values at the fourth block ($M = 0.160$, $SD = 0.02$), $t(21) = 2.73$, $p = 0.01$. However, the difference in the values within the first three blocks and the last two blocks did not reach significance (Figure 20). These results suggest that the boundary separation was narrower at the last two test blocks which was also consistent with the findings from Experiment 1. A decrease in the boundary separation parameter shows that participants responded faster, but less accurately at the last 60 test positions. This finding provides modest evidence for the attention hypothesis such that the decrease in attention may also contribute to the decrease in accuracy at the end of the test list. However, the pattern of this decrease does not suggest that attention is the only factor in OI.

In summary, the results from the DM analysis show that after studying a mixed strength list and being informed of the task used to encode the targets at the beginning of the test, participants may select different criterion for endorsing an item at test. The results from the DM fits show that the effect of a change in the criteria can be observed more prominently in the starting point parameter. A marginally significant difference in the drift rate parameter for foils as a function of strength also suggested a drift criterion shift in the current design. Further, the item-noise account of OI was supported by the decrease in the drift rates across test positions likewise in the previous experiments. Finally, the attention hypothesis was also supported as a contributing factor to the decrease in the accuracy especially in the last two blocks of the test lists.

Summary

In this experiment, we investigated whether an SBME would be observed after studying a list with mixed strength items and informing participants of the encoding task during the test. Second, we evaluated what type of decision processes produces the observed SBME. Item-noise models do not predict the SBME based on the differentiation mechanism. The memory evidence for a foil item does not change across list strength conditions because the foil items are matched to similar memory traces in terms of strength. However, the criterion shift account predicts lower false alarm rates for the strong test lists. The results from the accuracy data showed a tendency for an effect of strength on the false alarm rates. Moreover, the results from the DM analysis showed an effect of strength on the drift rates and the starting point parameters. Both of these parameters represent a different type of criteria in the DM and these results suggest that participants can adapt a different strategy for the strong and the weak test lists if they were informed about the test condition. However, when the size of these effects were compared to the size of the effects observed in Experiment 1, we conclude that both decision and memory processes were active in Experiment 1, where as only decision processes contribute to performance in Experiment 3. The purpose of Experiment 4 is to replicate the accuracy findings in Experiments 2 and 3.

Experiment 4

The goal of this experiment was to investigate how informing participants at the beginning of the test effects accuracy after studying a list of items with mixed strength. This experiment was a replication of Experiment 2 and Experiment 3 with minor differences in the methodology. The current experiment was less demanding compared to the previous experiments as the participants completed only one session with a lower number of study-test blocks and a lower number of trials in each study-test block. Consequently, there is an insufficient number of responses for a DM analysis of reaction time distributions. As the results from Experiment 2 showed, the SBME should not be observed if participants are not informed on the strength of the targets in the test list. On the other hand, when they are informed on the test list strength, as in Experiment 3, the SBME should be observed as participants set different criteria to endorse an item across the list strength conditions.

Methods

Participants

Eighty-two Syracuse University undergraduates participated in the experiment in exchange for course credits. Thirty four of the participants were female. Half of the participants were assigned to the condition in which they were informed on the type of the list that they will be tested on and the other half was assigned to the condition in which they were not informed. Participants who performed at low level of accuracy ($d' < 0.5$) excluded from the

analysis. The sample size for the subsequent analysis was 69: 36 for the informed condition, and 33 for the not-informed condition.

Materials

The words were selected from the word pool used in Experiments 1, 2 and 3.

Procedures and Design

Participants completed 4 blocks of study and test. Each study list consisted of 80 words. For each word, participants were given a levels-of-processing task. In each study list, for half of the words (40) participants were asked to make a semantic judgment ("Does the word have a pleasant meaning?") as a form of deep processing. For the other half (40), participants were asked to make an orthographic judgment ("Does the word contain the letter 'e'?") as a shallow processing task. The choice of the encoding task was random on each trial. The study trials were self-paced as participants responded by pressing the 'z' or '/?' keys on the keyboard and a 100 msec inter-stimulus interval followed the response. Two types of test lists were constructed as strong and weak. In the strong test lists, targets were the deeply encoded words, and in the weak test lists, targets were the shallowly encoded words. 40 new items were added as foils to the test list and as a result, the test list length was 80. In both test list conditions, targets and foils were presented in random order. In half of the study-test blocks (2), participants were tested on the strong lists and in the other half (2), participants were tested on the weak lists. The order of the test list was also randomized. The criterion placement was manipulated between-subjects by informing half of the participants on the strength of the list that they would be tested on and the other half were not informed on the test list strength. For each test trial, participants were asked to discriminate between targets and foils by a 'yes/no' recognition judgment. This experiment was a 2 (Strength) $\times$ 2 (Information) mixed design experiment, where the strength was a within-subjects factor and the information was a between-subjects factor.

Results and Discussions

A 2 (Strength) $\times$ 2 (Information) mixed factor repeated measures ANOVA was conducted to test the effects of strength and information on the hit rates and the false alarm rates.

Hit Rates A significant main effect of the list strength $F(1, 67) = 137.17$, $MSE = 1.603$, $p = .001$, ${\eta_p}^2 = 0.672$ was found as shown in top panel of Figure 21. Both groups of participants showed higher hit rates when they were tested on the strong targets ($M = 0.87$, $SD = 0.11$) compared to when they were tested on the weak targets ($M = 0.65$, $SD = 0.16$). Neither a main effect of information $F(1, 67) = 1.15$, $MSE = 0.014$, $p = .287$, ${\eta_p}^2 = 0.017$ nor an interaction between strength and information $F(1, 67) = 0.00$, $MSE \approx 0$, $p = .994$, ${\eta_p}^2 = 0.00$ were significant. These results suggest that the strength effect on the hit rates was robust and was not affected by information about the test targets.

False Alarm Rates The main effect of strength ($F(1, 67) = 4.485$, $MSE = 0.018$, $p = .038$, ${\eta_p}^2 = 0.063$) on the false alarm rates was found to be significant, as the mean false alarm rate was higher for the weak test list ($M = 0.21$, $SD = 0.16$) compared to the false alarm rate of the strong test list ($M = 0.18$, $SD = 0.14$). The main effect of information was failed to reach significance, $F(1, 67) = 2.865$, $MSE = 0.055$, $p = .095$, ${\eta_p}^2 = 0.041$. However, the interaction between the list strength and information was significant, $F(1, 67) = 10.125$, $MSE = 0.04$, $p = .002$, ${\eta_p}^2 = 0.131$.

Further analysis of contrasts show that the mean false alarm rate in the strong condition ($M = 0.195$, $SE = 0.17$) was significantly lower than the mean false alarm rate in the weak condition ($M = 0.25$, $SD = 0.19$), $t(35) = -3.62$, $p < .001$ only when the participants were informed about the encoding task of the targets on the test list. When participants were not informed, the mean false alarm rate of the strong list ($M = 0.17$, $SD = 0.10$) did not differ significantly from the mean false alarm rate of the weak test list

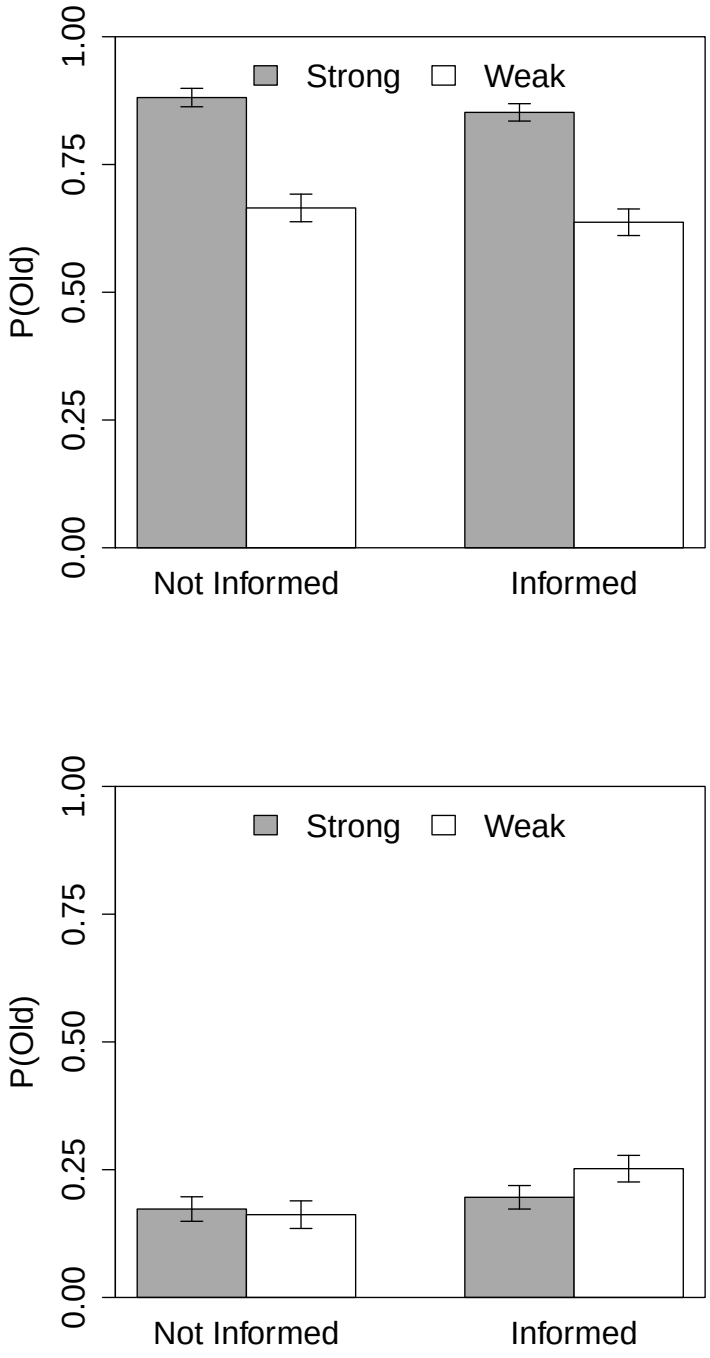

Figure 21: The hit rates and the false alarm rates as a function of list strength and information on the test list strength. The top panel presents the hit rates and the bottom panel presents the false alarm rates. Error bars represent one standard error above and below the mean.

($M = 0.16$, $SD = 0.09$), $t(32) = -0.789$, $p = .43$.

The result from the false alarm rates are important for two reasons. First, informing participants about the strength of the target words in a test list results in a change in criterion placement. When participants were told that they would be tested on only the weak targets, they shifted their criterion to be more liberal as they endorsed more foils while the mean false alarm rate for the strong test list was comparable to the false alarm rates observed in the not-informed condition (see the bottom panel of Figure 21).

Second, the current experiment has experimental parameters more in line with the Starns et al design, for example a shorter list length and a lower number of study-test blocks. We found that the strength effect on the false alarms became reliable and greater in magnitude than Experiment 3. Thus, these results show that participants can set different criterion for the test lists even when the information they receive is not as intuitive as the number of repetitions. However, they must be explicitly provided with information in order to change their criterion as they do not do so in the non-informed condition.

Summary

This experiment investigated in what way the test list strength can change the meta-cognitive decision processes of participants. Similar to Experiments 2 and 3, participants studied a mixed strength list and the list strength was manipulated at test by presenting either only the strong targets or the weak targets along with foils. One group of participants did not receive any information regarding the to-be-tested targets while the other group was informed. The results from the experiment showed that the group that was informed at the beginning of the test, adopted different criteria for the two different test conditions. Specifically, they selected a more lenient criterion for the weak test, resulting in higher false alarm rates for the weak foils. These findings show that participants can change their criterion when they are told to do so. However the existence

of a criterion shift is compatible with the REM model and provides no evidence against a differentiation process. Thus, overall the REM model is compatible with the finding of a SBME for pure study lists based on differentiation and a SBME for mixed-lists only when participants are given information that allows them to adjust the placement of the criteria.

General Discussion

One of the three goals of this project was to test whether a decrease in memory evidence or attention best explains OI. The results from the first three experiments show that the interference from testing has an extended effect on OI and that a decrease in attention also plays a role on the decrease in accuracy observed at the end of the test list. The second goal was to provide converging evidence for item-noise models as these models can explain both the SBME and OI with a single mechanism: the interference from the memory traces of other items. The results from the first experiment showed a SBME, OI and an interaction between the two, as predicted by REM. The final goal was to investigate whether the SBME could be observed after a mixed-strength study list when participants were not informed of the strength of the test list. The results from Experiment 2, 3 and 4 showed that a SBME is observed only when participants were informed and consequently, select different criteria for different test strength conditions. Each of these findings will be discussed in the following sections.

OI: Interference and Attention

To investigate the two explanations of OI, accuracy and reaction time data were analyzed as a function of test position in the first three experiments. The DM was used to further separate the effects of memory evidence and the attention processes on OI. A decrease in memory evidence as a function of test position manifested itself as a decrease in the drift rate parameters as the test progressed. This decrease was most prominent for all targets,

which is consistent with the faster rate of decrease in the hit rates as a function of test position. This pattern has been observed in all three experiments providing evidence for an increased interference that stems from active retrieval over the course of the test. The pattern of drift rates was less clear for foils across test positions as the decrease was non-monotonic. Additionally, we have observed a decrease in attention towards the end of the test list, which was estimated by the boundary separation parameter of the DM. The decrease in this parameter accounted for the faster and less accurate responses towards the end of the test list. However, the effect of test block on boundary separation was observed only at the last two test blocks which correspond to the last 60 test trials in all three experiments. This means that the boundary separation cannot explain the full pattern of decreasing accuracy over the course of the test. The results from the statistical models show a continuous decrease in accuracy from the start of the test list to the end. These findings suggest that although attention fluctuates during test and decreases at the end of a test session, OI is best explained by a decrease in the memory evidence caused by interference from earlier test trials, as REM predicts.

As discussed previously, the source of interference could be both item and context information, such as in REM, or solely context information (e.g., BCDMEM, Dennis & Humphreys, 2001). In the current study, simulations from REM provided evidence for the interference caused by other items in memory which were encoded at either study or test. However, these findings are mute to the context noise explanation. Although OI has not been implemented in the context-noise models yet, these models can also potentially explain the decrease in memory evidence towards the end of the test list by a drift in the reinstated context. In BCDMEM, reinstated context is matched to the retrieved context which contains all the contexts in which the item was encountered, including the study context. Thus, if one assumes that the reinstated context drifts over the course of the test list, then the match between the reinstated context and the retrieved context will be harmed. The results from Experiment 1 showed that the SBME and OI interact with each

other as the strong lists were less affected by active retrieval during test, producing a smaller OI effect. The context-reinstatement hypothesis would not explain why the reinstated context should drift differently in the two strength conditions. One could assume that the two strength conditions represent different contexts because list strength was manipulated through an orienting task (Alban & Kelley, 2012; Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005). In that case, BCDMEM would need to assume a faster drift in the shallow encoding context compared to the drift in the deep encoding context in order to explain the interaction between SBME and OI. Moreover, BCDMEM would also predict this interaction in Experiment 3 as participants were informed on the study context of the targets that they would be tested on. For example, participants were told that the test list would contain only the targets which they had made a semantic judgment on during study and therefore the reinstated context would reflect this information. In this test condition, the reinstated context would be expected to drift slower for the deep encoding condition, as in Experiment 1. However, the results from Experiment 3 showed that the decrease in the hit rates and the false alarm rates were at comparable levels across strong and weak test lists. To conclude, although these results are not contradictory for BCDMEM, additional assumptions and mechanisms would be required to explain the interaction between SBME and OI from the context-noise perspective.

SBME: Criterion shift and Differentiation

In this project, the two explanations of the SBME were investigated by manipulating item strength across and within study lists. The results from the first three experiments were simulated by REM and further analyzed with the DM. When the strength effect on the hit rates and the false alarm rates are compared across experiments, it could be suggested that both differentiation and criterion shift contribute to the SBME when strength is

manipulated at study across lists and only criterion shift explains the SBME observed after studying a mixed-strength list and being informed of the test condition.

The difference between the mean hit rates across strength conditions was consistent in all experiments of this study. The difference between the weak hit rate and the strong hit rate was 0.21 in Experiment 1, 0.23 in Experiment 2, 0.22 in Experiment 3 and 0.22 in Experiment 4. On the other hand, the difference in the mean false alarm rates across strong and weak test lists differed in each experiment: 0.10 in Experiment 1, 0.004 in Experiment 2, 0.03 in Experiment 3, 0.06 in the informed condition in Experiment 4 and 0.01 in the non-informed condition in Experiment 4. The false alarm rates which were produced in the pure-study list condition (Experiment 1) show a greater effect of strength compared to the mixed-study list conditions (Experiment 2-4) (cf. Starns et al., 2012). Thus, we conclude that the greater effect of the list strength in Experiment 1 stems from the additional role of differentiation in an across study list strength manipulation.

The simulations of REM in Experiment 2 showed that the parameters used to predict the hit rates and the false alarm rates observed in this study did not produce a SBME based on differentiation, even when the encoding-at-retrieval mechanism was implemented. According to the encoding-at-retrieval mechanism in REM, when an item receives an 'old' response, it updates the best matching trace and as a result, decreases the probability of a match between memory and subsequent foils. Thus, since more items are endorsed in the strong lists because of greater memory evidence for the targets, the false alarm rate is expected to decrease towards the end of the test list. When an item is rejected, it is added as a new trace, increasing the noise in memory. The subsequently tested foils would more likely produce a match value that exceeds the criterion. As the memory evidence is lower in the weak test lists, more targets will be rejected and consequently, add a new trace in memory. This mechanism would be expected to increase the false alarm rates over the course of a weak test list. Although neither the empirical findings nor the model simulations conveyed such a pattern, one can question whether this

effect of updating or adding a new trace can be observed in shorter time scales. Could this explain the differentiation at retrieval mechanism and the SBME observed after mixed-study lists? For example, after three consecutive 'old' responses, memory would be updated and the foil that is presented right after these responses would match poorly with the traces in memory. Then, we would expect to see a decrease in the false alarm rates as a function of consecutive 'old' responses. Similarly, after 3 consecutive 'new' responses, three new traces will be added to memory and therefore, increase the search set and noise in memory. In that case, we would expect to observe higher false alarm rates for the foils that were followed by these responses. Figure 22 plots the false alarm rates collapsed across test strength conditions as a function of the number of previous consecutive responses in Experiment 2 and 3. The white points represent the consecutive 'old' responses and the black points represent the consecutive 'new' responses. Contrary to the model predictions, the false alarm rates increased as a function of consecutive 'old' responses and decreased as a function of 'new' responses. Recently, Malmberg and Annis (2012) showed that sequential dependencies exist in recognition memory such that participants are more likely to give an 'old' response to a foil, if the previous response was a false alarm rather than a correct rejection. Malmberg and Annis (2012) also showed that these dependencies exist even for the responses with two trials apart. Our findings further reflect a build up of sequential dependencies. As the number of previous 'old' responses increases, the probability of an 'old' response for a foil item increases. Likewise, as the number of prior 'new' responses increases, the probability to of a 'new' response increases. In this regard, it is possible that the effect of sequential dependencies might be exceeding the effect of differentiation in the false alarm rates over three trials and preventing us from observing the effect of differentiation at retrieval over a short time scale.

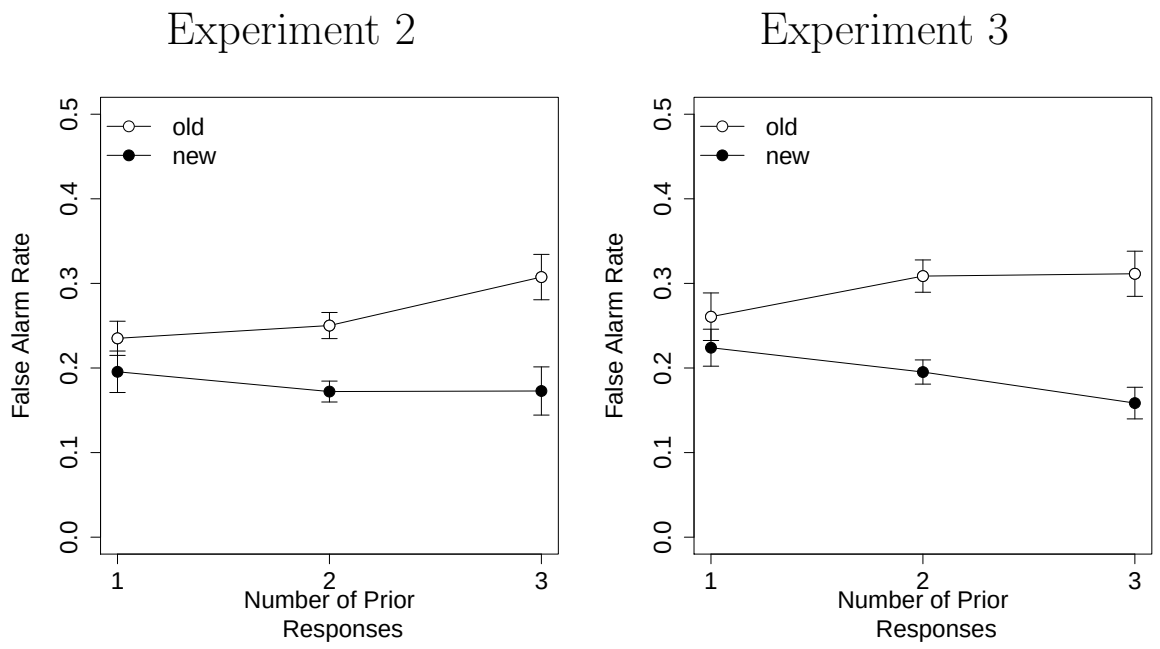

Figure 22: Sequential dependencies in the false alarm rates. Error bars are within-subjects 95% CI.

Does criterion setting have to be separate from memory?

One of the goals of this project was to investigate whether the criterion shift account or the encoding at retrieval mechanism is the reason for observing a SBME after participants study a list of items with mixed strength. Results from Experiment 2-4 showed that the SBME was observed only when participants were informed of the strength of the targets which they will be tested on along with foils. This provides evidence for the criterion shift. The DM fits to Experiment 2 and 3 also displayed a very small decrease in the drift rates (in absolute value) for the strong foils compared to the weak foils as previously reported by Starns et al. (2012). This suggests that the drift rate parameter is sensitive to criterion manipulations through the zero point drift rate parameter: the drift criterion (Ratcliff, 1978, 1985; Ratcliff & McKoon, 2008; Ratcliff et al., 1999). In this regard, the DM also has problems in separating the decision processes from the strength of evidence in any detection task. The problem arises mainly from the uncertainty of the exact placement of

the criterion.

In the SDT literature, memory evidence and criterion have been assumed to be independent of each other and as a result, these two processes were assumed to be separable. According to SDT, however, the criterion placement along the evidence scale can not be determined, only the relative placement of the criterion. Recent models have been developed to overcome this criterion placement problem by suggesting a dynamic mechanism of evidence accumulation over time to produce subjective target and foil distributions (e.g., Cox & Shiffrin, 2012; Turner, Van Zandt, & Brown, 2011). Thus, instead of setting a criterion (e.g., drift criterion in the DM), these models propose that the criterion placement depends on evidence, and provide a theoretical account of how these subjective distributions are formed over the course of a response to a single test item (Cox & Shiffrin, 2012) or over the number of trials in a test list (Turner et al., 2011). In the future, OI and the SBME should be tested from a dynamic criterion setting account and, these models could be evaluated with the paradigm used in this project.

Conclusions

This project provided converging evidence on the role of interference from other items in memory. Results from the experiments show that interference caused by both study and test items can be explained by an item-noise model, REM. The results from the DM demonstrated that memory and decision processes are intertwined in recognition memory, both contributing to performance.

Appendix A

Parameters from Individual Fits

Across Experiment Comparisons of Drift Rates

The drift rate parameter estimates obtained from the DM#2 that was fit to each individual in Experiments 1-3 was compared by using a 2 (Strength) $\times$ 3 (Experiment) mixed ANOVA. The ANOVA results for the target drift rates revealed a significant main effect of strength, $F(1, 73) = 185.98$, $p < .001$. The mean target drift rates were significantly greater in the strong condition ($M = 0.20$, $SD = 0.12$) compared to the mean target drift rates in the weak condition ($M = 0.04$, $SD = 0.16$). The main effect of experiment did not reach significance nor did the interaction between the strength condition and experiment.

The ANOVA results for the foil drift rates showed a significant main effect of strength, $F(1, 73) = 43.07$, $p < .001$. The mean drift rate for the strong foils ($M = -0.21$, $SD = 0.10$) were significantly greater than the mean drift rate for the weak foils ($M = -0.16$, $SD = 0.10$) in absolute values. However, a significant interaction between the strength condition and experiment ($F(2, 73) = 10.12$, $p < .001$) showed that a significant strength effect is observed only in Experiments 1 ($v_{SF} = -0.23$, $v_{WF} = -0.14$) and 3 ($v_{SF} = -0.17$, $v_{WF} = -0.14$), $t(139) = 12.57$, $p < .001$; $t(109) = 3.13$, $p < .01$, respectively. The strength effect on the foil drift rates did not reach significance in Experiment 2

($v_{SF} = -0.21$, $v_{WF} = -0.20$). Further contrasts showed that the strength effect was greater in Experiment 1 compared to the strength effect in Experiment 3, $t(222) = 5.41$, $p < .001$. Finally, ANOVA did not reveal a significant main effect of experiment on the foil drift rates.

In addition to the analysis on the drift rate estimates of targets and foils, the difference between the drift rates of targets and foils were analyzed across experiments. The difference between the mean drift rate of targets and the mean drift rate of foils is an estimate of the sensitivity index (d') in the signal detection theory. As defined in the DM, drift rate is an estimate of the mean rate of accumulation which is assumed to be normally distributed with a standard deviation of η. Since η was fixed across conditions in the models that were fit to individual data, the mean drift rates for targets and foils were standardized similar to the d' in the signal detection theory. Thus, the difference between the mean target drift rates and the mean foil drift rates can be considered as an estimate of d' for drift rates ($v_{Target} - v_{Foil}$). In previous DM applications, this difference between the drift rates were fixed across conditions to measure the change in the drift criterion (e.g., Criss, 2010; Ratcliff & McKoon, 2008). If the drift d' is fixed across conditions but the drift rate values change, that difference in the drift rates would be due to a change in the criterion. However, in the current analysis we will use this measure to compare the effect of strength on the rate of evidence accumulation across experiments. A 2 (Strength) $\times$ 3 (Experiment) mixed ANOVA was conducted on the drift rate d' values obtained from individual fits of DM#2. The results revealed a main effect of strength, $F(1, 73) = 200.54$, $p < .001$. The drift d' was greater in the strong condition compared to the weak condition in all three experiments (Figure A.1). Neither experiment nor the interaction between the two factors were significant. These findings suggest that the strength effect on the accuracy component of the drift rates did not differ across experiments as the item-noise models would predict. Thus, the differences observed in the drift rates across experiments could be due to a change in the drift criterion across experiments.

Tables A.1 - A.4 show the parameters from the DM#2 that was fit to individuals in

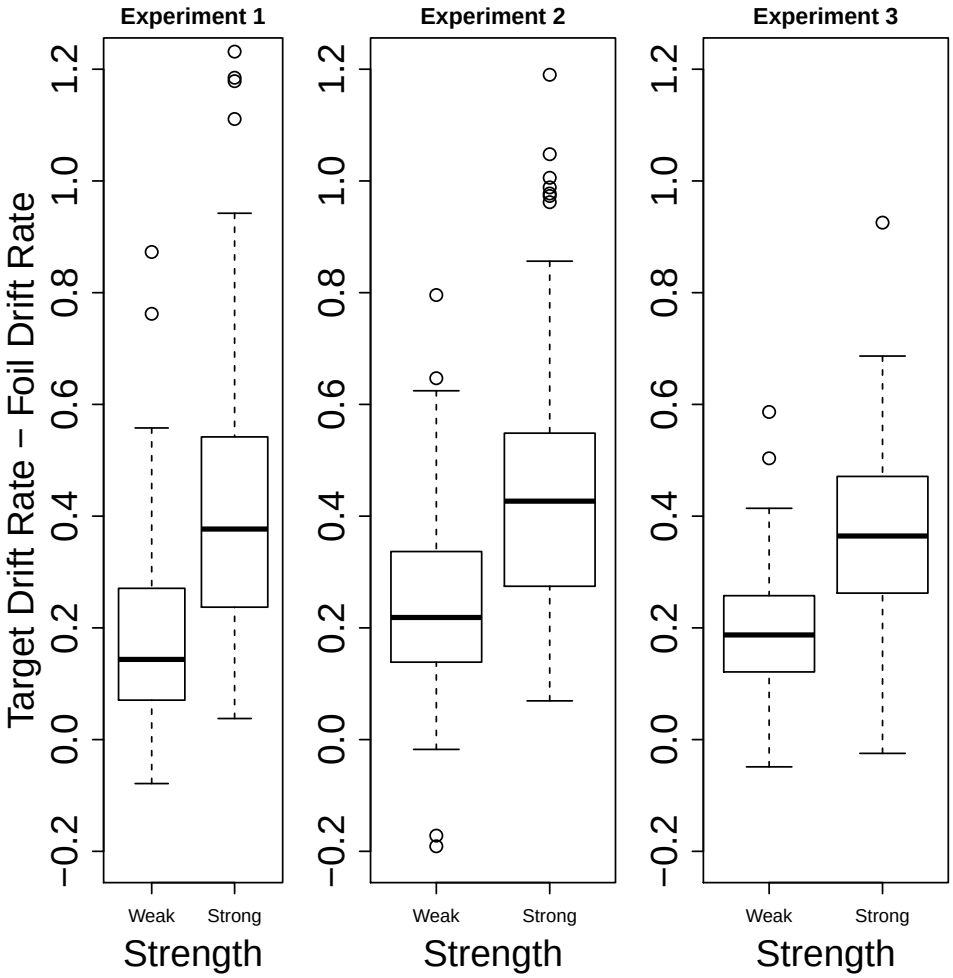

Figure A.1: d' of the drift rates across the strength condition and experiments

Experiment 1, Tables A.5 - A.8 display the parameters from the DM#2 that was fit to individuals in Experiment 2 and Tables A.9 - A.12 show the parameters from the DM#2 that was fit to individuals in Experiment 3. Tables A.13 - A.15 show the parameters averaged over Tables A.1 - A.12.

	Weak Targets					Weak Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.2327	0.5323	0.5598	0.1093	0.0866	-0.3251	-0.2299	-0.3132	-0.2448	-0.3127
2	0.1017	-0.0239	-0.0404	-0.0695	-0.1449	-0.0099	-0.0581	-0.1263	-0.0370	-0.2014
3	0.0285	0.0450	0.0042	-0.0385	-0.1449	-0.2007	-0.0958	-0.1728	-0.1041	-0.1882
4	0.1882	0.0157	0.0149	0.0035	-0.0686	0.0020	-0.0135	-0.0345	-0.0043	-0.0916
5	0.2163	0.1988	0.1876	0.1070	0.0265	-0.1269	-0.0011	-0.1057	-0.0856	-0.1026
6	0.2209	0.1452	0.1205	0.1453	0.0933	-0.2855	-0.2163	-0.1583	-0.2719	-0.1912
7	0.2089	0.1582	0.1658	0.0071	0.0125	-0.2159	-0.1227	-0.2016	-0.2297	-0.3200
8	-0.0270	0.0296	0.0179	0.0673	-0.1328	-0.3082	-0.2547	-0.2239	-0.2587	-0.2638
9	0.1571	0.1253	0.1657	0.1052	0.0437	-0.1546	-0.0819	-0.1491	-0.0622	-0.0711
10	-0.0011	0.0119	0.0190	0.0241	0.0653	0.0200	-0.0893	-0.0839	-0.0414	-0.0543
11	0.0960	-0.0438	-0.0186	-0.1591	-0.1269	-0.1253	-0.1616	-0.2020	-0.2984	-0.2232
12	0.1891	0.0281	0.0178	-0.0911	-0.1294	-0.1160	-0.1185	-0.2195	-0.1831	-0.2210
13	0.0452	0.0650	0.0119	0.0153	-0.0225	-0.0948	-0.0578	-0.0538	-0.0937	-0.0669
14	0.1938	0.1963	-0.0023	0.1498	0.0091	-0.0261	-0.0721	-0.0583	-0.0029	-0.1380
15	0.3682	0.1326	0.1097	0.0271	-0.1632	-0.1360	0.0190	-0.1634	-0.2035	-0.2099
16	0.1246	-0.0449	-0.0237	-0.0662	-0.1672	-0.0788	-0.1166	-0.1264	-0.1622	-0.1363
17	0.1790	0.1773	-0.1494	-0.2232	-0.2857	-0.2007	-0.2104	-0.3139	-0.2554	-0.2756
18	0.0055	0.0191	0.0524	0.0171	-0.0645	-0.1567	-0.0865	-0.0794	-0.1590	-0.1151
19	0.1276	0.0438	0.0342	0.0354	-0.1150	-0.0019	-0.0257	-0.0364	-0.1027	-0.0963
20	0.2986	0.2529	0.1088	-0.0067	-0.0292	-0.0751	-0.1817	-0.0217	-0.1478	-0.1781
21	0.3049	0.2046	0.0406	0.0368	0.0121	-0.1451	-0.0067	-0.1064	-0.1728	-0.2110
22	0.1340	0.1204	0.0172	-0.2061	-0.1781	-0.2016	-0.2105	-0.2616	-0.3372	-0.2355
23	0.0330	0.0141	-0.0329	-0.0794	-0.1517	-0.1142	-0.1446	-0.1855	-0.1801	-0.2050
24	-0.0309	-0.0083	-0.0123	-0.1010	-0.1795	-0.0470	-0.0491	-0.0553	-0.0778	-0.1008
25	0.1678	0.0008	0.1206	-0.0873	-0.0272	0.0282	-0.1527	-0.1390	-0.1795	-0.0822
26	0.2780	0.2351	-0.0052	-0.0965	-0.1845	-0.1086	-0.1876	-0.1989	-0.3194	-0.2290
27	0.1052	0.0649	0.0448	-0.0069	-0.0554	-0.0407	-0.0756	-0.0471	-0.0775	-0.1202
28	-0.0358	-0.1297	-0.1609	-0.0863	-0.1528	-0.3271	-0.1284	-0.1309	-0.2110	-0.1635

Table A.1: Drift rate estimates for the weak condition for each individual from DM#2 in Experiment 1.

	Strong Targets					Strong Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.6188	0.7166	0.5602	0.4508	0.5425	-0.5660	-0.5149	-0.5507	-0.3815	-0.6362
2	0.1136	0.0545	0.0276	0.0226	-0.0575	-0.1199	-0.1606	-0.1184	-0.1640	-0.1770
3	0.2222	0.1010	0.0912	-0.0037	0.0014	-0.3653	-0.2601	-0.2048	-0.2141	-0.1864
4	0.3344	0.2534	0.1431	0.1512	0.1487	-0.1216	-0.1755	-0.2022	-0.1977	-0.1619
5	0.3723	0.3331	0.2507	0.1564	0.1274	-0.1468	0.0076	-0.1597	-0.1482	-0.1273
6	0.4593	0.3386	0.2352	0.2522	-0.0344	-0.3792	-0.4328	-0.3651	-0.2764	-0.2700
7	0.5172	0.4226	0.3823	0.4306	0.2472	-0.3660	-0.2778	-0.3447	-0.1903	-0.2143
8	0.3790	0.2725	0.2225	0.0881	0.1634	-0.4375	-0.3360	-0.3832	-0.3588	-0.2809
9	0.2867	0.2357	0.3469	0.1672	0.3744	-0.1606	-0.2262	-0.1452	-0.1889	-0.2034
10	0.1566	0.1207	0.1367	-0.0045	0.2946	-0.1211	-0.0755	-0.0941	-0.1853	-0.2452
11	0.1301	0.1538	-0.0672	0.0960	0.0572	-0.3556	-0.2886	-0.3105	-0.4245	-0.2075
12	0.1629	0.2258	0.0738	0.0424	0.0354	-0.1940	-0.1507	-0.1676	-0.2003	-0.2428
13	0.0828	0.0760	0.0795	0.0050	0.0276	-0.1577	-0.1304	-0.1185	-0.1812	-0.1288
14	0.3271	0.2731	0.3212	0.2291	0.1785	-0.2319	-0.1825	-0.1967	-0.1808	-0.2918
15	0.3040	0.2044	0.2840	0.1965	0.1410	-0.2799	-0.2351	-0.2871	-0.1757	-0.2364
16	0.1266	0.0728	0.0884	0.0003	0.0023	-0.1027	-0.1356	-0.1461	-0.1931	-0.1246
17	0.2636	0.0431	0.2562	-0.1153	0.0485	-0.3272	-0.3059	-0.3899	-0.2591	-0.3841
18	0.2737	0.2933	0.3031	0.1677	0.1879	-0.2673	-0.2299	-0.1773	-0.2258	-0.1492
19	0.0971	0.1162	0.0400	0.0684	0.0536	-0.1414	-0.2104	-0.1538	-0.1454	-0.1548
20	0.5151	0.3755	0.4525	0.5560	0.3118	-0.2057	-0.2435	-0.2359	-0.3862	-0.2467
21	0.3494	0.2524	0.2093	0.0342	0.1493	-0.2232	-0.1899	-0.0930	-0.2050	-0.1695
22	0.3195	0.3951	0.1446	0.1891	0.1323	-0.2789	-0.3192	-0.3526	-0.3619	-0.3371
23	0.2410	0.2670	0.1752	0.1376	0.0430	-0.1575	-0.2155	-0.2327	-0.2162	-0.2224
24	0.0165	-0.0102	-0.0467	-0.0784	-0.0596	-0.0603	-0.0479	-0.1201	-0.1285	-0.1213
25	0.2884	0.1668	0.1263	0.0046	0.0152	-0.1067	-0.1138	-0.2107	-0.1772	-0.2146
26	0.4817	0.3840	0.2595	0.2768	0.2390	-0.2353	-0.2482	-0.3366	-0.2660	-0.2682
27	0.1235	0.1740	0.2681	0.1313	0.1168	-0.1030	-0.1033	-0.1224	-0.1072	-0.1241
28	0.0057	-0.0241	-0.0748	-0.0495	-0.0658	-0.2470	-0.2734	-0.1872	-0.1899	-0.1767

Table A.2: Drift rate estimates for the strong condition for each individual from DM#2 in Experiment 1.

	Weak Starting Point					Strong Starting Point				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.0772	0.0716	0.0752	0.0590	0.0630	0.1016	0.0942	0.0989	0.0777	0.0829
2	0.0947	0.0994	0.0963	0.0891	0.0842	0.1095	0.1149	0.1114	0.1030	0.0974
3	0.0810	0.0847	0.0806	0.0732	0.0724	0.0921	0.0962	0.0916	0.0832	0.0822
4	0.0574	0.0547	0.0545	0.0580	0.0523	0.0631	0.0601	0.0598	0.0637	0.0575
5	0.0957	0.1012	0.1026	0.0966	0.0985	0.1018	0.1077	0.1092	0.1028	0.1048
6	0.0541	0.0499	0.0510	0.0499	0.0500	0.0732	0.0675	0.0690	0.0676	0.0676
7	0.0961	0.0937	0.0879	0.0812	0.0838	0.1271	0.1240	0.1164	0.1074	0.1109
8	0.0782	0.0802	0.0791	0.0747	0.0745	0.0864	0.0885	0.0874	0.0826	0.0823
9	0.0447	0.0460	0.0453	0.0461	0.0471	0.0553	0.0569	0.0561	0.0571	0.0583
10	0.0949	0.0999	0.0964	0.0890	0.0909	0.0990	0.1042	0.1005	0.0929	0.0948
11	0.0756	0.0688	0.0664	0.0654	0.0590	0.0716	0.0652	0.0629	0.0619	0.0559
12	0.0779	0.0747	0.0736	0.0684	0.0721	0.0923	0.0885	0.0872	0.0810	0.0854
13	0.0933	0.0921	0.0861	0.0922	0.0872	0.1010	0.0998	0.0933	0.0998	0.0944
14	0.1038	0.1001	0.1031	0.0979	0.1082	0.1102	0.1063	0.1094	0.1039	0.1148
15	0.1134	0.1093	0.1150	0.1106	0.1094	0.1106	0.1066	0.1121	0.1078	0.1067
16	0.0732	0.0756	0.0757	0.0724	0.0648	0.0598	0.0617	0.0618	0.0591	0.0529
17	0.0611	0.0581	0.0568	0.0600	0.0566	0.0904	0.0860	0.0841	0.0889	0.0838
18	0.1081	0.1137	0.1004	0.1077	0.1076	0.1131	0.1189	0.1049	0.1126	0.1125
19	0.0514	0.0571	0.0556	0.0539	0.0533	0.0669	0.0742	0.0722	0.0700	0.0693
20	0.1044	0.1067	0.1075	0.1094	0.1027	0.1141	0.1166	0.1175	0.1196	0.1122
21	0.0793	0.0740	0.0711	0.0658	0.0738	0.0711	0.0663	0.0638	0.0590	0.0662
22	0.0949	0.0972	0.0952	0.1009	0.0922	0.1059	0.1085	0.1062	0.1126	0.1028
23	0.0686	0.0722	0.0719	0.0706	0.0714	0.0764	0.0804	0.0802	0.0787	0.0795
24	0.0484	0.0489	0.0506	0.0516	0.0481	0.0522	0.0527	0.0546	0.0556	0.0519
25	0.0707	0.0688	0.0665	0.0603	0.0588	0.0811	0.0789	0.0763	0.0692	0.0675
26	0.1038	0.1014	0.1045	0.1020	0.0936	0.1193	0.1165	0.1201	0.1172	0.1076
27	0.0926	0.0994	0.1061	0.0871	0.1029	0.0980	0.1052	0.1123	0.0922	0.1089
28	0.0535	0.0559	0.0492	0.0487	0.0466	0.0602	0.0629	0.0553	0.0548	0.0524

Table A.3: Starting point estimates for each individual from DM#2 in Experiment 1. This table presents the starting point parameter values which are changing relative to the boundary separation parameter value. Since the relative starting point was fixed across test blocks, the ratio of the starting point value to the boundary separation value is also fixed across test blocks.

	Boundary Separation					T_{er}	s_T	s_z	η
Test Block	1	2	3	4	5				
Participant									
1	0.1884	0.1747	0.1834	0.1440	0.1536	0.7616	0.4825	0.1084	0.2504
2	0.2036	0.2136	0.2071	0.1915	0.1811	0.6193	0.2086	0.1160	0.4318
3	0.1875	0.1961	0.1866	0.1695	0.1675	0.5152	0.2889	0.1437	0.3537
4	0.1311	0.1250	0.1244	0.1325	0.1195	0.5057	0.1532	0.1036	0.3455
5	0.1775	0.1878	0.1904	0.1792	0.1828	0.4693	0.2021	0.0000	0.0000
6	0.1458	0.1346	0.1374	0.1347	0.1348	0.5304	0.3695	0.0988	0.2003
7	0.2322	0.2265	0.2125	0.1962	0.2025	0.5838	0.3492	0.1594	0.2757
8	0.1936	0.1985	0.1959	0.1851	0.1845	0.6129	0.3334	0.1480	0.2094
9	0.0980	0.1009	0.0995	0.1012	0.1034	0.5082	0.1040	0.0000	0.1407
10	0.1992	0.2097	0.2024	0.1869	0.1909	0.4386	0.1851	0.1770	0.4643
11	0.1973	0.1796	0.1733	0.1707	0.1539	0.4770	0.3078	0.1107	0.2633
12	0.1625	0.1558	0.1534	0.1425	0.1504	0.5167	0.1551	0.0000	0.2118
13	0.1693	0.1672	0.1563	0.1674	0.1582	0.5776	0.1207	0.0000	0.2232
14	0.2152	0.2077	0.2138	0.2030	0.2243	0.6208	0.3040	0.1948	0.4895
15	0.2199	0.2119	0.2229	0.2144	0.2122	0.5994	0.3110	0.1655	0.2012
16	0.1585	0.1637	0.1638	0.1568	0.1402	0.5328	0.1308	0.1048	0.3791
17	0.1954	0.1858	0.1817	0.1921	0.1811	0.5147	0.4118	0.1122	0.3526
18	0.1914	0.2013	0.1776	0.1907	0.1904	0.5023	0.2288	0.0000	0.0124
19	0.1343	0.1491	0.1452	0.1407	0.1393	0.5080	0.1785	0.1018	0.3871
20	0.1999	0.2042	0.2058	0.2096	0.1966	0.5645	0.3247	0.1551	0.2658
21	0.1866	0.1741	0.1674	0.1549	0.1737	0.4918	0.3360	0.1170	0.2451
22	0.2002	0.2050	0.2007	0.2128	0.1944	0.6023	0.3427	0.1786	0.2190
23	0.1389	0.1462	0.1457	0.1429	0.1445	0.5636	0.1824	0.0431	0.1658
24	0.1686	0.1704	0.1765	0.1798	0.1677	0.4373	0.0000	0.0071	0.1958
25	0.1605	0.1561	0.1509	0.1370	0.1335	0.5153	0.2839	0.1166	0.2184
26	0.2071	0.2022	0.2085	0.2033	0.1867	0.5995	0.3173	0.1465	0.2272
27	0.1691	0.1815	0.1937	0.1590	0.1878	0.5275	0.1500	0.0000	0.1718
28	0.1832	0.1914	0.1685	0.1668	0.1594	0.4448	0.2097	0.0921	0.3323

Table A.4: Boundary separation and other fixed parameter estimates for each individual from DM#2 in Experiment 1.

	Weak Targets					Weak Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.3780	0.0170	0.1578	0.1196	-0.0495	-0.2690	-0.2502	-0.1986	-0.2673	-0.4699
2	0.1585	0.1101	-0.0157	-0.1369	-0.1071	-0.2787	-0.1896	-0.1795	-0.2811	-0.2031
3	0.3458	0.3891	0.2032	0.1587	-0.0635	-0.4500	-0.2353	-0.3229	-0.1780	-0.3483
4	0.1004	0.1271	0.0627	-0.0072	-0.2037	-0.4361	-0.1824	-0.2558	-0.2742	-0.3451
5	0.0750	0.0990	0.0824	-0.0040	-0.0065	-0.0928	-0.0823	-0.1080	-0.1285	-0.0873
6	-0.0181	-0.0024	-0.0489	-0.0663	-0.0644	-0.1373	-0.0989	-0.0758	-0.0986	-0.0983
7	-0.0164	0.0274	-0.0585	-0.0685	-0.0672	-0.2044	-0.1993	-0.2821	-0.2665	-0.2082
8	0.3312	0.1761	0.1722	0.1827	0.0063	-0.2750	-0.2394	-0.2123	-0.2731	-0.2118
9	0.1453	-0.0068	-0.0134	-0.1805	-0.1022	-0.2411	-0.3307	-0.3697	-0.2994	-0.3119
10	0.2109	0.0421	-0.0686	-0.0520	-0.0870	-0.1889	-0.2102	-0.2701	-0.2578	-0.3013
11	-0.0442	0.0544	0.0483	0.0863	-0.0159	-0.2289	-0.2222	-0.2707	-0.2780	-0.4149
12	0.0833	0.0296	0.0992	0.0523	0.0159	-0.3034	-0.2786	-0.2744	-0.3368	-0.3534
13	0.1567	-0.0345	-0.0313	0.0815	-0.1125	-0.2166	-0.2059	-0.2067	-0.1825	-0.3473
14	0.1053	0.0309	0.0196	-0.0320	-0.0904	-0.1451	-0.0802	-0.0941	-0.1480	-0.3299
15	0.0722	0.0634	0.1011	-0.0102	-0.0796	-0.2180	-0.1794	-0.1898	-0.2501	-0.3053
16	-0.0492	-0.0868	-0.0884	-0.1028	-0.1456	-0.2220	-0.1855	-0.1721	-0.1917	-0.2343
17	-0.1103	-0.0598	-0.1077	-0.0707	-0.1008	-0.2392	-0.1142	-0.1304	-0.3991	-0.1388
18	0.0052	0.0256	0.0625	0.0083	0.0015	-0.0759	-0.1233	-0.0505	-0.0101	-0.0262
19	-0.0318	-0.0326	-0.0722	-0.0816	-0.0922	-0.1739	-0.1595	-0.1668	-0.2730	-0.2987
20	0.2607	0.2103	0.2264	0.2235	0.2171	-0.0280	-0.0397	0.0101	0.1556	-0.0164
21	-0.1537	0.0401	0.0007	0.0639	0.0999	-0.1504	-0.1322	-0.2185	-0.2126	-0.2026
22	0.0960	0.0486	0.0302	-0.1030	-0.4239	-0.1089	-0.1735	-0.1129	-0.2417	-0.2329
23	0.2184	0.1446	-0.0129	0.0692	0.0467	-0.2094	-0.1975	-0.2652	-0.2970	-0.3466
24	0.2482	0.3202	0.1916	0.0275	-0.0865	-0.2462	0.1084	-0.2358	0.0450	0.0853
25	0.4188	0.2177	0.1846	0.1821	-0.0182	-0.0757	-0.1204	-0.2324	-0.1516	-0.1720
26	0.0564	0.0208	0.0171	0.0014	-0.0804	-0.1390	-0.1229	-0.1317	-0.1519	-0.2142

Table A.5: Drift rate estimates for the weak condition for each individual from DM#2 in Experiment 2.

	Strong Targets					Strong Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.2892	0.2651	0.4064	0.1351	0.2435	-0.3086	-0.2462	-0.2715	-0.2905	-0.2545
2	0.3350	0.3030	0.2024	0.1560	0.1600	-0.2041	-0.2455	-0.2873	-0.2907	-0.2534
3	0.7171	0.4962	0.5716	0.3738	0.2715	-0.4730	-0.2390	-0.4016	-0.3033	-0.3476
4	0.5401	0.4916	0.2358	0.1598	0.1865	-0.2849	-0.2718	-0.2893	-0.4557	-0.4506
5	0.3600	0.3334	0.2776	0.1921	0.2862	-0.1665	-0.1301	-0.1516	-0.1123	-0.1313
6	0.1786	0.1650	0.0691	0.0997	0.0394	-0.1883	-0.0725	-0.1317	-0.1265	-0.1196
7	0.3028	0.2077	0.1671	0.1545	0.1708	-0.2219	-0.1854	-0.1796	-0.2166	-0.1567
8	0.2416	0.1973	0.2024	0.0896	0.0816	-0.2570	-0.2491	-0.2392	-0.2796	-0.4531
9	0.3816	0.2254	0.1440	0.1052	0.1341	-0.1293	-0.3122	-0.1946	-0.3303	-0.1901
10	0.7968	0.7122	0.7663	0.3215	0.4164	-0.2511	-0.2654	-0.2394	-0.3768	-0.4400
11	0.3365	0.3524	0.1416	0.1880	0.1874	-0.2862	-0.3271	-0.2865	-0.2348	-0.2199
12	0.4853	0.4382	0.4767	0.3635	0.2047	-0.3180	-0.3399	-0.3268	-0.3831	-0.4559
13	0.3506	0.4758	0.3365	0.2974	0.2494	-0.3191	-0.5126	-0.2807	-0.2416	-0.2290
14	0.1845	0.1565	0.1097	0.0081	0.0783	-0.0708	-0.1001	-0.1096	-0.1739	-0.1453
15	0.3368	0.3080	0.3064	0.1911	0.1041	-0.0778	-0.2261	-0.2391	-0.4423	-0.3493
16	0.1053	0.0236	0.0448	-0.0000	-0.0337	-0.1214	-0.1590	-0.1061	-0.1364	-0.1640
17	0.1037	0.2359	0.0536	0.0439	0.0376	-0.1455	-0.0985	-0.1586	-0.2458	-0.2093
18	0.1936	0.1283	0.1017	0.1382	0.0778	-0.1416	-0.0485	0.0324	-0.0052	-0.0227
19	0.0946	0.1325	0.1485	-0.0189	0.0478	-0.1799	-0.1888	-0.3009	-0.2349	-0.2106
20	0.4375	0.5829	0.5542	0.3530	0.2804	0.0401	-0.0576	0.0138	-0.0310	-0.0517
21	0.3175	0.2031	0.2031	0.1554	0.1327	-0.2072	-0.1647	-0.2655	-0.1946	-0.2452
22	0.3708	0.3274	0.1490	0.1536	0.0683	-0.0782	-0.1235	-0.1707	-0.1401	-0.1907
23	0.1897	0.0029	-0.0014	0.0193	-0.0010	-0.4158	-0.2280	-0.2410	-0.2563	-0.2344
24	0.2575	0.5110	0.3512	0.3126	0.2329	-0.2801	-0.4510	-0.2224	-0.2437	-0.2703
25	0.2244	0.3636	-0.0377	-0.0030	-0.0269	-0.1253	-0.1275	-0.1407	-0.1537	-0.1549
26	0.1897	0.1814	0.1684	0.0253	0.1033	-0.1315	-0.1076	-0.1212	-0.1021	-0.1213

Table A.6: Drift rate estimates for the strong condition for each individual from DM#2 in Experiment 2.

	Weak Starting Point					Strong Starting Point				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.0397	0.0397	0.0403	0.0405	0.0407	0.0513	0.0513	0.0521	0.0523	0.0526
2	0.1147	0.1221	0.1286	0.1391	0.1124	0.0994	0.1058	0.1114	0.1206	0.0974
3	0.0786	0.0751	0.0887	0.0836	0.0714	0.0859	0.0821	0.0970	0.0914	0.0780
4	0.0881	0.0933	0.0951	0.0920	0.0897	0.1080	0.1144	0.1167	0.1128	0.1100
5	0.0989	0.1104	0.1114	0.1025	0.1091	0.0986	0.1101	0.1111	0.1022	0.1088
6	0.0733	0.0881	0.0860	0.0845	0.0844	0.0811	0.0974	0.0951	0.0935	0.0933
7	0.0775	0.0758	0.0755	0.0822	0.0792	0.0727	0.0712	0.0709	0.0772	0.0744
8	0.1102	0.1176	0.1099	0.1128	0.1111	0.1100	0.1174	0.1097	0.1126	0.1109
9	0.0655	0.0736	0.0620	0.0644	0.0669	0.0660	0.0741	0.0625	0.0648	0.0674
10	0.0860	0.0887	0.0903	0.0875	0.0934	0.0907	0.0936	0.0953	0.0923	0.0985
11	0.0567	0.0558	0.0584	0.0582	0.0586	0.0611	0.0601	0.0629	0.0627	0.0631
12	0.0703	0.0719	0.0703	0.0719	0.0742	0.0630	0.0644	0.0629	0.0643	0.0664
13	0.0716	0.0665	0.0695	0.0695	0.0618	0.0945	0.0879	0.0918	0.0918	0.0817
14	0.0506	0.0505	0.0486	0.0433	0.0470	0.0446	0.0445	0.0427	0.0382	0.0414
15	0.0644	0.0688	0.0682	0.0765	0.0732	0.0755	0.0807	0.0800	0.0897	0.0859
16	0.0828	0.0809	0.0845	0.0839	0.0795	0.0773	0.0755	0.0788	0.0782	0.0741
17	0.0529	0.0486	0.0486	0.0486	0.0507	0.0523	0.0481	0.0481	0.0481	0.0502
18	0.0859	0.0837	0.0774	0.0698	0.0682	0.0995	0.0969	0.0897	0.0809	0.0790
19	0.0950	0.0877	0.0858	0.0840	0.0773	0.1004	0.0927	0.0907	0.0888	0.0817
20	0.1308	0.1319	0.1345	0.1420	0.1412	0.1293	0.1304	0.1330	0.1404	0.1396
21	0.0966	0.1046	0.1071	0.1042	0.1018	0.1049	0.1136	0.1164	0.1132	0.1105
22	0.0920	0.0962	0.0995	0.0942	0.0986	0.0993	0.1039	0.1074	0.1017	0.1065
23	0.0668	0.0629	0.0609	0.0618	0.0598	0.0693	0.0653	0.0632	0.0641	0.0621
24	0.0582	0.0672	0.0644	0.0630	0.0709	0.0637	0.0736	0.0705	0.0690	0.0777
25	0.0876	0.1047	0.1021	0.1006	0.1084	0.0856	0.1022	0.0997	0.0982	0.1058
26	0.1068	0.1123	0.1147	0.1082	0.1231	0.0901	0.0947	0.0968	0.0913	0.1038

Table A.7: Starting point estimates for each individual from DM#2 in Experiment 2. This table presents the starting point parameter values which are changing relative to the boundary separation parameter value. Since the relative starting point was fixed across test blocks, the ratio of the starting point value to the boundary separation value is also fixed across test blocks.

	Boundary Separation					T_{er}	s_T	s_z	η
Test Block	1	2	3	4	5				
Participant									
1	0.1266	0.1266	0.1287	0.1292	0.1298	0.4780	0.3491	0.0784	0.1862
2	0.1970	0.2098	0.2208	0.2389	0.1929	0.5960	0.3060	0.1367	0.2808
3	0.1791	0.1711	0.2021	0.1904	0.1625	0.5831	0.3042	0.1417	0.2530
4	0.1894	0.2006	0.2046	0.1978	0.1929	0.6073	0.3834	0.1436	0.1991
5	0.1835	0.2048	0.2068	0.1903	0.2025	0.5140	0.1711	0.0000	0.1354
6	0.1726	0.2074	0.2025	0.1990	0.1987	0.5375	0.1060	0.0748	0.2252
7	0.1542	0.1509	0.1502	0.1636	0.1576	0.5584	0.2006	0.1408	0.2129
8	0.2097	0.2239	0.2092	0.2147	0.2114	0.5703	0.3550	0.1824	0.1216
9	0.1313	0.1475	0.1243	0.1290	0.1341	0.6765	0.3111	0.1227	0.2559
10	0.1621	0.1672	0.1703	0.1649	0.1760	0.5483	0.4160	0.1401	0.2803
11	0.1285	0.1265	0.1323	0.1318	0.1327	0.5722	0.2581	0.1107	0.2106
12	0.1592	0.1628	0.1591	0.1627	0.1680	0.4968	0.3525	0.1248	0.2635
13	0.1864	0.1733	0.1810	0.1811	0.1610	0.4971	0.2883	0.1226	0.3164
14	0.1154	0.1152	0.1107	0.0988	0.1072	0.4446	0.1219	0.0000	0.2860
15	0.1394	0.1490	0.1477	0.1657	0.1586	0.5158	0.2834	0.1267	0.2687
16	0.1560	0.1523	0.1591	0.1580	0.1497	0.4199	0.0755	0.1394	0.3273
17	0.1388	0.1274	0.1275	0.1274	0.1331	0.5394	0.1839	0.0951	0.2797
18	0.1769	0.1723	0.1594	0.1438	0.1405	0.5017	0.1499	0.1220	0.5431
19	0.2277	0.2103	0.2055	0.2013	0.1852	0.5863	0.1777	0.1536	0.2058
20	0.2200	0.2219	0.2263	0.2389	0.2375	0.5882	0.2960	0.1774	0.2393
21	0.1943	0.2104	0.2155	0.2097	0.2047	0.5815	0.2603	0.1768	0.1864
22	0.1654	0.1729	0.1789	0.1693	0.1773	0.4813	0.1636	0.0000	0.2310
23	0.1817	0.1711	0.1656	0.1680	0.1627	0.4692	0.2893	0.1187	0.2021
24	0.1291	0.1491	0.1428	0.1397	0.1573	0.4981	0.4349	0.1154	0.1671
25	0.1619	0.1934	0.1887	0.1858	0.2003	0.6386	0.3700	0.1066	0.2255
26	0.1926	0.2026	0.2070	0.1952	0.2220	0.5958	0.1612	0.0635	0.3051

Table A.8: Boundary separation and other fixed parameter estimates for each individual from DM#2 in Experiment 2.

	Weak Targets					Weak Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	-0.0275	-0.0432	-0.0319	-0.0938	-0.0641	-0.0887	-0.1289	-0.1363	-0.0953	-0.1521
2	0.2583	0.2611	0.1085	-0.0490	0.0691	-0.2453	-0.0677	-0.0198	-0.0002	-0.1198
3	0.1533	0.0967	0.0160	0.0957	0.0873	-0.2118	-0.0445	-0.1889	-0.1500	-0.1100
4	0.2544	0.1435	0.1519	0.0790	0.1293	-0.1190	-0.0366	-0.1300	-0.1710	-0.1025
5	0.1168	0.0256	0.1151	0.0916	0.1087	-0.1808	-0.0967	-0.0476	-0.1584	-0.1272
6	0.1545	0.2308	0.1004	0.0767	0.0384	-0.1644	-0.1833	-0.1803	-0.1458	-0.1548
7	0.0019	0.0181	0.0542	-0.0403	0.0336	-0.1298	-0.0977	-0.0833	-0.1358	-0.0415
8	0.1252	0.1156	0.0929	0.0088	0.0153	-0.1620	-0.1096	-0.1880	-0.1598	-0.1461
9	0.3108	0.0725	0.0641	0.1148	-0.0557	-0.2757	-0.2225	-0.2109	-0.1532	-0.4324
10	0.0095	-0.0381	0.0317	-0.1528	-0.2133	-0.1787	-0.2043	-0.2487	-0.1820	-0.2114
11	0.0128	-0.0082	-0.0489	-0.0522	-0.1123	-0.2571	-0.2049	-0.2786	-0.1810	-0.3409
12	-0.0033	0.0166	-0.0133	-0.0258	0.0305	-0.0848	-0.0476	-0.1008	-0.0844	-0.0792
13	0.1939	0.2959	0.1967	0.0943	0.0951	0.0186	0.0380	0.0460	-0.0480	-0.0159
14	0.0081	0.0003	0.0799	-0.0207	0.0272	-0.1288	-0.0899	-0.1571	-0.1568	-0.1373
15	0.0573	0.0287	0.0483	-0.0059	-0.0551	-0.0949	-0.0504	-0.0971	-0.0849	-0.1182
16	0.0882	0.1106	-0.0735	0.0232	-0.0357	-0.3095	-0.1954	-0.2320	-0.2189	-0.3140
17	0.1058	0.0661	0.0490	0.0325	0.0308	-0.1518	-0.1203	-0.0917	-0.1268	-0.0795
18	0.0161	0.0315	0.0355	-0.1067	-0.0168	-0.2453	-0.2155	-0.2324	-0.2723	-0.2359
19	0.0867	0.0097	-0.0897	-0.2244	-0.1754	-0.1354	-0.2009	-0.2885	-0.3675	-0.2965
20	-0.0131	0.0331	0.0189	-0.0821	0.0463	-0.1254	-0.1858	-0.1094	-0.1542	-0.1317
21	0.2196	0.1563	0.0769	0.0448	-0.0881	-0.1365	-0.0769	-0.2230	-0.1917	-0.1105
22	0.2198	0.2631	0.1058	-0.0715	0.1149	-0.0030	-0.0107	0.1004	-0.1772	-0.1426

Table A.9: Drift rate estimates for the weak condition for each individual from DM#2 in Experiment 3.

	Strong Targets					Strong Foils				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	-0.0738	-0.0593	-0.1115	-0.0660	-0.1172	-0.2274	-0.2013	-0.2179	-0.1564	-0.2186
2	0.6635	0.2674	0.5147	0.4157	0.3650	-0.2620	-0.2620	-0.1719	-0.0882	-0.2686
3	0.2536	0.2988	0.2004	0.2335	0.1453	-0.1324	-0.1531	-0.1559	-0.0924	-0.1952
4	0.2856	0.2354	0.2337	0.2019	0.2220	-0.2556	-0.2980	-0.2646	-0.1626	-0.1946
5	0.3886	0.3423	0.2993	0.3445	0.2647	-0.1632	-0.1337	-0.1810	-0.1459	-0.0888
6	0.3919	0.3005	0.4054	0.2157	0.2017	-0.1867	-0.1706	-0.2155	-0.2005	-0.2246
7	0.1733	0.1989	0.1234	-0.0148	-0.0141	-0.1911	-0.2230	-0.2282	-0.2226	-0.1845
8	0.2726	0.1534	0.1370	0.0903	0.0338	-0.1527	-0.1575	-0.1276	-0.1624	-0.1766
9	0.4420	0.2166	0.2520	0.1660	0.1256	-0.2352	-0.1679	-0.3903	-0.1783	-0.2345
10	0.3164	0.3241	0.0283	-0.0329	0.0062	0.0791	-0.1550	-0.2321	-0.1621	-0.2513
11	0.0833	0.2153	0.1146	0.0093	0.0030	-0.1789	-0.1557	-0.2189	-0.2002	-0.2471
12	0.0245	0.1496	0.0281	0.0053	-0.0029	-0.2232	-0.2079	-0.1889	-0.2060	-0.2300
13	0.4694	0.5921	0.3549	0.4298	0.2701	-0.1103	-0.0851	-0.0797	-0.0867	-0.0496
14	0.2176	0.2616	0.1858	0.0824	0.0746	-0.0941	-0.1300	-0.1798	-0.2312	-0.1687
15	-0.0258	0.0226	-0.0178	-0.0441	-0.0650	-0.1469	-0.0839	-0.0868	-0.1273	-0.0404
16	0.2575	0.3245	0.2093	0.2392	0.1722	-0.2151	-0.2290	-0.2539	-0.2812	-0.3423
17	0.3315	0.2732	0.1888	0.2410	0.1838	-0.1559	-0.1166	-0.1377	-0.1125	-0.1674
18	0.2606	0.2125	0.2626	0.2618	0.0155	-0.1040	-0.2365	-0.1974	-0.2468	-0.2143
19	0.1828	0.2292	0.0800	0.1037	0.0332	-0.1842	-0.1607	-0.2596	-0.2125	-0.2248
20	0.1931	0.1140	0.1187	0.0433	0.1239	-0.1446	-0.1725	-0.2054	-0.2200	-0.2484
21	0.3088	0.3863	0.2955	0.2749	0.1765	-0.1248	-0.1216	-0.1086	-0.1463	-0.1751
22	0.4804	0.3874	0.3433	0.2523	0.3312	0.0862	0.0102	0.0163	-0.0205	-0.0869

Table A.10: Drift rate estimates for the strong condition for each individual from DM#2 in Experiment 3.

	Weak Starting Point					Strong Starting Point				
Test Block	1	2	3	4	5	1	2	3	4	5
Participant										
1	0.0679	0.0666	0.0800	0.0708	0.0707	0.0493	0.0483	0.0581	0.0514	0.0513
2	0.0811	0.0809	0.0844	0.0809	0.0793	0.0857	0.0856	0.0892	0.0855	0.0839
3	0.0616	0.0509	0.0600	0.0558	0.0598	0.0801	0.0662	0.0780	0.0726	0.0778
4	0.0692	0.0815	0.0759	0.0761	0.0730	0.0778	0.0917	0.0854	0.0855	0.0821
5	0.0807	0.0773	0.0793	0.0835	0.0660	0.0760	0.0728	0.0747	0.0786	0.0621
6	0.1056	0.1082	0.1041	0.0953	0.0951	0.1119	0.1147	0.1103	0.1010	0.1007
7	0.0811	0.0771	0.0768	0.0781	0.0689	0.1113	0.1058	0.1053	0.1072	0.0945
8	0.1035	0.0947	0.0993	0.1055	0.1030	0.1154	0.1056	0.1107	0.1177	0.1149
9	0.0574	0.0594	0.0574	0.0543	0.0544	0.0776	0.0803	0.0776	0.0734	0.0736
10	0.0661	0.0711	0.0634	0.0580	0.0580	0.0761	0.0818	0.0730	0.0668	0.0668
11	0.0584	0.0518	0.0516	0.0512	0.0493	0.0737	0.0654	0.0651	0.0646	0.0622
12	0.0688	0.0605	0.0584	0.0588	0.0548	0.0677	0.0595	0.0574	0.0579	0.0540
13	0.0504	0.0538	0.0499	0.0500	0.0508	0.0629	0.0672	0.0622	0.0623	0.0634
14	0.0646	0.0652	0.0663	0.0623	0.0646	0.0787	0.0794	0.0807	0.0758	0.0787
15	0.0897	0.0813	0.0746	0.0739	0.0669	0.0757	0.0686	0.0630	0.0624	0.0564
16	0.0790	0.0906	0.0877	0.0843	0.0907	0.1002	0.1148	0.1111	0.1069	0.1150
17	0.0997	0.1006	0.0943	0.0934	0.1019	0.0881	0.0889	0.0834	0.0826	0.0901
18	0.0607	0.0596	0.0610	0.0603	0.0591	0.0594	0.0583	0.0597	0.0590	0.0579
19	0.0466	0.0469	0.0435	0.0418	0.0412	0.0695	0.0699	0.0648	0.0623	0.0614
20	0.0612	0.0608	0.0653	0.0638	0.0592	0.0795	0.0789	0.0849	0.0829	0.0769
21	0.0712	0.0719	0.0708	0.0658	0.0657	0.0774	0.0783	0.0770	0.0716	0.0715
22	0.0697	0.0743	0.0640	0.0622	0.0610	0.0764	0.0814	0.0701	0.0681	0.0668

Table A.11: Starting point estimates for each individual from DM#2 in Experiment 3. This table presents the starting point parameter values which are changing relative to the boundary separation parameter value. Since the relative starting point was fixed across test blocks, the ratio of the starting point value to the boundary separation value is also fixed across test blocks.

	Boundary Separation					T_{er}	s_T	s_z	η
Test Block	1	2	3	4	5				
Participant									
1	0.1644	0.1613	0.1938	0.1715	0.1713	0.6297	0.1158	0.0956	0.3539
2	0.1678	0.1675	0.1747	0.1674	0.1642	0.5177	0.3087	0.1576	0.2558
3	0.1529	0.1264	0.1489	0.1385	0.1486	0.4503	0.0815	0.1008	0.3668
4	0.1314	0.1549	0.1443	0.1445	0.1387	0.6045	0.2454	0.1059	0.2447
5	0.1558	0.1491	0.1531	0.1611	0.1273	0.5617	0.2447	0.1216	0.3197
6	0.2034	0.2085	0.2007	0.1836	0.1832	0.5666	0.3367	0.1546	0.2503
7	0.2171	0.2064	0.2055	0.2091	0.1843	0.5147	0.2110	0.1367	0.3848
8	0.2234	0.2043	0.2142	0.2277	0.2223	0.5755	0.2546	0.1850	0.3301
9	0.1661	0.1720	0.1662	0.1571	0.1576	0.4846	0.2614	0.1076	0.1429
10	0.1987	0.2138	0.1906	0.1745	0.1745	0.4778	0.3294	0.1151	0.3469
11	0.1812	0.1608	0.1601	0.1587	0.1529	0.4831	0.1654	0.0975	0.3066
12	0.1501	0.1319	0.1273	0.1282	0.1196	0.5004	0.1457	0.1069	0.3665
13	0.1078	0.1151	0.1066	0.1068	0.1087	0.5066	0.1645	0.0595	0.2172
14	0.1815	0.1831	0.1860	0.1748	0.1815	0.4853	0.2324	0.0000	0.1727
15	0.1748	0.1585	0.1455	0.1441	0.1304	0.2854	0.0573	0.0910	0.0806
16	0.1699	0.1947	0.1884	0.1812	0.1950	0.5396	0.2986	0.1383	0.1949
17	0.1755	0.1771	0.1661	0.1644	0.1794	0.5298	0.1542	0.1410	0.3461
18	0.1529	0.1501	0.1536	0.1519	0.1489	0.5443	0.3092	0.1147	0.2355
19	0.1556	0.1566	0.1451	0.1396	0.1374	0.4928	0.1753	0.0813	0.3445
20	0.1661	0.1649	0.1774	0.1732	0.1607	0.5690	0.2297	0.1174	0.2471
21	0.1687	0.1705	0.1678	0.1559	0.1558	0.4790	0.2797	0.1304	0.1825
22	0.1421	0.1514	0.1305	0.1267	0.1242	0.4739	0.4636	0.1139	0.3472

Table A.12: Boundary separation and other fixed parameter estimates for each individual from DM#2 in Experiment 3.

Individual Parameter Values (DM #1)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.151	0.279	-0.123	-0.248	0.179	
2	0.101	0.237	-0.120	-0.222	0.179	
3	0.045	0.212	-0.151	-0.245	0.179	
4	-0.016	0.145	-0.177	-0.253	0.179	
5	-0.073	0.137	-0.190	-0.248	0.179	
Fixed Parameters	z/a_W 0.45	z/a_S 0.50	η 0.281	s_z 0.109	T_{er} 0.545	s_T 0.271

Individual Parameter Values (DM #2)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.139	0.270	-0.127	-0.231	0.179	
2	0.092	0.224	-0.112	-0.217	0.179	
3	0.048	0.189	-0.142	-0.228	0.176	
4	-0.016	0.129	-0.161	-0.226	0.170	
5	-0.078	0.122	-0.172	-0.225	0.168	
Fixed Parameters	z/a_W 0.45	z/a_S 0.50	η 0.258	s_z 0.096	T_{er} 0.540	s_T 0.249

Table A.13: The average parameter values from DM#1 and DM#2 fit to individual data (Experiment 1).

Individual Parameter Values (DM #1)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.132	0.382	−0.233	-0.225	0.177	
2	0.085	0.349	−0.191	-0.223	0.177	
3	0.034	0.251	−0.212	-0.213	0.177	
4	−0.004	0.173	−0.229	-0.228	0.177	
5	−0.069	0.155	−0.268	-0.239	0.177	
Fixed Parameters	z/a_W 0.48	z/a_S 0.50	η 0.255	s_z 0.125	T_{er} 0.547	s_T 0.294

Individual Parameter Values (DM #2)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.117	0.320	−0.206	-0.205	0.168	
2	0.078	0.301	−0.163	-0.210	0.174	
3	0.044	0.236	−0.192	-0.204	0.174	
4	0.013	0.154	−0.202	-0.230	0.172	
5	−0.062	0.143	−0.236	-0.233	0.171	
Fixed Parameters	z/a_W 0.48	z/a_S 0.50	η 0.246	s_z 0.112	T_{er} 0.542	s_T 0.260

Table A.14: The average parameter values from DM#1 and DM#2 fit to individual data (Experiment 2).

Individual Parameter Values (DM #1)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.117	0.295	-0.171	-0.173	0.168	
2	0.102	0.271	-0.137	-0.183	0.168	
3	0.049	0.200	-0.153	-0.197	0.168	
4	-0.013	0.183	-0.170	-0.192	0.168	
5	-0.001	0.144	-0.193	-0.209	0.168	
Fixed Parameters	z/a_W 0.43	z/a_S 0.48	η 0.283	s_z 0.127	T_{er} 0.517	s_T 0.252

Individual Parameter Values (DM #2)

Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.107	0.268	-0.155	-0.151	0.168	
2	0.086	0.247	-0.116	-0.164	0.167	
3	0.049	0.193	-0.140	-0.186	0.166	
4	-0.012	0.157	-0.155	-0.166	0.161	
5	0.001	0.116	-0.163	-0.192	0.158	
Fixed Parameters	z/a_W 0.43	z/a_S 0.48	η 0.274	s_z 0.112	T_{er} 0.512	s_T 0.230

Table A.15: The average parameter values from DM#1 and DM#2 fit to individual data (Experiment 3).

Appendix B

Additional DM fits

Drift Criterion or Starting Point

In Experiment 2 and 3, a mixed-study list paradigm was used in order to test the criterion shift account of the SBME. Since the list strength is manipulated only at test in a mixed-study list paradigm, the foils at both strong and weak test conditions are expected to produce similar levels of false alarm rates by the differentiation account of the SBME. This is due to the similar levels of the study list strength across the test conditions. In this case, a SBME observed in a mixed-study paradigm should be due to the criterion shift account. In a previous study, Starns et al. (2012) extended this paradigm to the DM showing that drift rates are also affected by the test list strength condition, and argued that the effect observed in the drift rates are due to the changes in the drift criterion. They also demonstrated that the starting point changes across the strong and the weak test lists and the data are best explained when both of these parameters are allowed to vary. In the main text, we have presented the model in which both the starting point and the drift rate parameters were allowed to vary across test list strength conditions. Although AICc prefers the model presented in the text, here we further assess the list strength effect on the drift criterion and the starting point parameters separately by fixing either of the parameters

while relaxing the other one.

Experiment 2

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)
1	1115	22	210944	0.00	211136	0.00
2	1147	26	210984	0.00	211211	0.00
3	1075	24	210913	1.00	211123	1.00
4	1108	29	210956	0.00	211209	0.00

Experiment 3

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)
1	1428	22	178785	1.00	178973	1.00
2	1485	26	178859	0.00	179082	0.00
3	1452	24	178798	0.00	179003	0.00
4	1540	29	178913	0.00	179161	0.00

Table B.1: The diffusion model comparisons for the drift criterion and the starting point parameter models. The χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion with finite sample correction,w_i(AICc) is the Akaike weights which represents the probability of the model being the best model, BIC is the Bayesian Information Criterion, w_i(BIC) is the Schwarz weights which represents the probability of the model being the best model according to the BIC values.

Four additional diffusion models were fit to the data from Experiment 2 and 3. In the first model, the drift rates were fixed across the foil strength conditions. The drift rate of the strong foils were estimated to be equal to the drift rate of the weak foils. The starting point parameter was allowed to vary across strength conditions and to account for OI, the drift rates were also allowed to vary across test positions. Thus, there were a total of 22 free parameters in the model (10 from v_T, 5 from v_F, 2 from z, a, T_{er}, s_T, s_z, η). The first model was fit to test the starting point, whereas the second model was fit to test the drift criterion hypothesis.

In the second model, the starting point parameter was fixed across strength conditions and instead, the drift rates of the foils were relaxed across test list strength.

This model yielded 26 free parameters (10 from v_T, 10 from v_F, z, a, T_{er}, s_T, s_z, η).

In the third and the fourth model, the attention hypothesis of OI was added to the first two models by relaxing boundary separation across test positions. These models were fit to make a direct comparison between the model preferred in the main text and the starting point-only or the drift rate-only models presented here. In the third model, the response bias was fixed to the z/a obtained from the first model to prevent the starting point from changing as a function of test position. In the third model, the number of free parameters increased to 26 (10 from v_T, 5 from v_F, 5 from a, T_{er}, s_T, s_z, η).

In the last model, the starting point parameter was fixed and the drift rates were allowed to vary across 20 conditions as in the second model, and the boundary separation parameter was let to vary across test positions. The number of free parameters was 29 (10 from v_T, 10 from v_F, z, 5 from a, T_{er}, s_T, s_z, η).

Table B.1 presents the model fit statistics from Experiment 2 and 3. In Experiment 2, AICc and BIC favored the third model, suggesting that the strength effect observed in a mixed-study list paradigm is explained better by a change in the starting point parameter, rather than a change in the drift rates across foils. In Experiment 3, AICc and BIC preferred the first model among the models that were discussed here. The first model also suggests that the shift in the criterion is explained better with the starting point parameter, and in addition, does not require the boundary separation parameter to vary across test positions. The important finding here is that both preferred models favor the starting point as the main contributing factor in the criterion shift in the SBME observed in a mixed-study list paradigm (c.f. Starns et al., 2012). This is also consistent with the relatively slight differences in the mean drift rates across the test strength conditions when both the drift rate and the starting point parameters were allowed to vary in Experiment 2 and 3, respectively. The parameters of the preferred models are presented in Table B.2. Notably, in both experiments, AICc prefers the model that was presented in the main text over the models presented here. This suggests that both the drift rate and the starting

point parameter are required to change across strength conditions (as the best fitting model presented in Starns et al. (2012)) along with a boundary separation changing across test positions in order to best explain the variance in the data.

Group Parameter Values (DM #3 in Exp 2)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.118	0.354	−0.243	-0.243	0.176	
2	0.091	0.316	−0.229	-0.243	0.181	
3	0.036	0.263	−0.235	-0.235	0.181	
4	0.003	0.167	−0.253	-0.253	0.177	
5	−0.065	0.126	−0.275	-0.275	0.175	
Fixed Parameters	z/a_W 0.47	z/a_S 0.49	η 0.207	s_z 0.155	T_{er} 0.579	s_T 0.292

Group Parameter Values (DM #1 in Exp 3)						
Test Block	v_{WT}	v_{ST}	v_{WF}	v_{SF}	a	
1	0.106	0.253	−0.203	-0.203	0.163	
2	0.080	0.259	−0.164	-0.164	0.163	
3	0.027	0.182	−0.193	-0.193	0.163	
4	0.026	0.155	−0.207	-0.207	0.163	
5	0.025	0.146	−0.239	-0.239	0.163	
Fixed Parameters	z/a_W 0.43	z/a_S 0.48	η 0.255	s_z 0.141	T_{er} 0.548	s_T 0.261

Table B.2: The group parameter values of the DM tested the starting point parameter across the strength conditions in Experiment 2 and 3. In the model that was preferred in Experiment 2 among the four tested models in Appendix B, the boundary separation was fixed across test blocks.

Starting Point Parameter Across Test Block

Two additional diffusion models were fit to the data from the first three experiments to investigate whether the response bias changes across test positions. In order to do so, the starting point parameter was also allowed to vary in addition to the best fitting models

presented in the body of this project. The model fit statistics from these additional models suggest that the response bias changes in addition to the drift rates and the boundary separation parameters. None of the models discussed in this project predict a response bias change as a function of the test position to account for OI. REM, as an example, did not need such a manipulation in the *criterion* parameter to account for the empirical findings observed in the first three experiments regarding OI. As OI refers to a decrease in accuracy, the change in the response bias as a function of test position was not critical for the purposes of this project. However, these results are presented here for future investigations.

Experiment 1

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)
1	1615	39	227955	1.00	228298	1.00
2	1875	23	228189	0.00	228392	0.00

Experiment 2

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)
1	988	39	210856	1.00	211196	1.00
2	1347	23	211169	0.00	211369	0.00

Experiment 3

DM #	χ^2	df	AICc	w_i(AICc)	BIC	w_i(BIC)
1	1293	39	178708	1.00	179042	1.00
2	1392	23	178756	0.00	178953	0.00

Table B.3: The diffusion model comparisons for the models that assess the starting point parameter as a function of test position in the first three experiments. The χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion with finite sample correction, w_i(AICc) is the Akaike weights which represents the probability of the model being the best model, BIC is the Bayesian Information Criterion, w_i(BIC) is the Schwarz weights which represents the probability of the model being the best model according to the BIC values.

In the first model, the drift rate, boundary separation and the starting point parameters were adjusted freely as a function of test position. In order to account for the SBME, the drift rate and the starting point parameters were allowed to vary as a function of strength in accord with the best fitting models presented in the body of this project. After adding the test position conditions on the starting point parameter, the number of free parameters in the model increased to 39 (20 from v, 10 from z, 5 from a, T_{er}, s_T, s_z, η).

The second model assessed whether the starting point and boundary separation can account for OI without an adjustment in drift rate. Hence, this was a reduced model of the first one, and a comparison between the two would test the need for a parameter that moderates the memory evidence as a function test position. In this model, the boundary separation and the starting point parameters were allowed to vary across test positions while the drift rate was fixed. To account for the SBME, both the drift rate and the boundary separation parameters were allowed to vary across the strength condition. The total number of parameters in this model was 23 (4 from v, 10 from z, 5 from a, T_{er}, s_T, s_z, η).

Table B.3 displays the model fit statistics for the model comparisons of the three experiments. In all three experiments, AICc and BIC favors the first model, suggesting that the drift rates are required to change across test positions to best explain the reaction time and accuracy data in all three experiments. The item-noise explanation was further supported by these findings. AICc favors the first model tested here over the best fitting models presented in the main text for all three experiments. This comparison between the first model and the best fitting model presented in the body of this project demonstrates that, in addition to drift rate and boundary separation, an inclusion of the starting point parameter best explains the data observed in these experiments.

Table B.4, Table B.5 and Table B.6 present the parameter estimates of the first model from the three experiments respectively. The starting point parameter was found to decrease consistently as a function of test position in all three experiments. The decrease in

		Experiment 1			
	Test Block	v_T	v_F	z/a	a
Weak	1	0.089	-0.113	0.492	0.159
	2	0.042	-0.117	0.483	0.161
	3	0.042	-0.110	0.436	0.156
	4	−0.006	-0.139	0.415	0.154
	5	−0.022	-0.123	0.385	0.154
Strong	1	0.199	-0.193	0.502	0.159
	2	0.175	-0.175	0.492	0.161
	3	0.145	-0.199	0.489	0.156
	4	0.127	-0.197	0.459	0.154
	5	0.108	-0.200	0.462	0.154
Fixed Parameters		η 0.221	s_z 0.118	T_{er} 0.572	s_T 0.189

Table B.4: The diffusion model parameters of the starting point model for Experiment 1.

		Experiment 2			
	Test Block	v_T	v_F	z/a	a
Weak	1	0.104	-0.247	0.498	0.171
	2	0.062	-0.223	0.492	0.178
	3	0.027	-0.202	0.485	0.175
	4	−0.023	-0.226	0.464	0.175
	5	−0.017	-0.221	0.437	0.173
Strong	1	0.313	-0.247	0.507	0.171
	2	0.260	-0.266	0.535	0.178
	3	0.253	-0.231	0.492	0.175
	4	0.175	-0.245	0.494	0.175
	5	0.164	-0.258	0.473	0.173
Fixed Parameters		η 0.203	s_z 0.150	T_{er} 0.577	s_T 0.278

Table B.5: The diffusion model parameters of the starting point model for Experiment 2.

the starting point parameter shows that the response bias also changes over the course of the test since participants become more stringent for endorsing a test item towards the end

		Experiment 3			
	Test Block	v_T	v_F	z/a	a
Weak	1	0.059	-0.121	0.472	0.143
	2	0.058	-0.076	0.445	0.143
	3	0.059	-0.083	0.401	0.142
	4	0.009	-0.099	0.403	0.138
	5	0.036	-0.076	0.363	0.139
Strong	1	0.156	-0.135	0.500	0.143
	2	0.153	-0.132	0.513	0.143
	3	0.123	-0.141	0.486	0.142
	4	0.119	-0.124	0.459	0.138
	5	0.099	-0.131	0.458	0.139
Fixed Parameters		η 0.225	s_z 0.100	T_{er} 0.527	s_T 0.121

Table B.6: The diffusion model parameters of the starting point model for Experiment 3.

of the test list. These findings, coupled with the decrease in the drift rates, suggest that as task difficulty increases towards the end of the test list, participants tend to adjust their response bias accordingly, showing a tendency to reject the test item (Benjamin & Bawa, 2004; Hirshman, 1995). This finding contradicts the previous findings by Verde and Rotello (2007), showing that the response bias does not change over the course of the test list, even when the test difficulty increased by study-strength manipulation. In their experiments (1-4), participants studied a mixed-study list and were later test on pure-strength lists which were blocked. While the strong test lists were presented always in the first block, the weak test list was presented in the second block. Thus, the first test block was easier as the participants had better memory for strong items, and the second test block was considered to be difficult compared to the first block. Their results did not show a difference in the false alarms across the test blocks, and they concluded that the difficulty of the test list does not alone cause a change in the response bias. We should also note that, participants were not informed on the strength of the test list which they would be presented. However,

in Experiment 4, when Verde and Rotello (2007) gave feedback on participants' accuracy at each test trial, participants shifted their criterion to be more stringent in the second (difficult) block. They suggested that feedback influenced the shift in the criterion in a similar fashion to informing participants on the test list strength. It can be argued that these two manipulations give the participants the opportunity to adjust their criterion. Additionally, the results from our experiment showed that response bias could change over the course of the test list, even when participants were not given any feedback.

Foil Drift Rates Fixed Across Test Block

To further test the item-noise account of OI, an additional diffusion model was fit to data from Experiment 1, 2 and 3. In the body of the document, an ANOVA was applied to test the effect of test block and the strength condition on the drift rate parameters that were obtained from the individual fits (DM#2). The results from the ANOVA suggested that the test block did not have a significant effect on the drift rates of foils. As a result, an additional model that fixed the drift rates of the foils across test blocks was fit to group data (DM#3). There were 21 free parameters (10 from v_T, 2 from v_F, 5 from a, T_{er}, s_z, s_T, and η). The starting point parameters were fed into the model from the response bias estimates of DM#1 that was fit in the body of the dissertation. The current model (DM#3) was compared to the best fitting model (DM#2) in Table B.7. In Experiment 1, both AICc and BIC preferred DM#2 in which the drift rates for both targets and foils were let to vary across test block. Thus, the DM comparisons suggest that the drift rates decrease as a function of test position contrary to the results from ANOVA. Model fit statistics from Experiment 2 showed that AICc preferred DM#2 and BIC preferred DM#3. These results suggested an inconclusive model selection. Finally, in Experment 3, both AICc and BIC preferred the reduces model (DM#3) and suggested that fixing the drift rates of foils across test blocks did not harm the model fit. Similar to ANOVA results,

DM#3 suggests that the decrease in the drift rate as a function of test position was not significant for foils in Experiment 3. Table B.8 presents the parameter estimates from DM#3 in three experiments.

Experiment 1

DM #	χ^2	df	AICc	BIC
2	1675	29	228028	228283
3	1814	21	228116	228301

Experiment 2

DM #	χ^2	df	AICc	BIC
2	1061	29	210909	211162
3	1092	21	210933	211116

Experiment 2

DM #	χ^2	df	AICc	BIC
2	1429	29	178806	179054
3	1312	21	178689	178689

Table B.7: The diffusion model comparisons for the foil drift rates. The χ^2 values are from the fits to the group data. AICc is the Akaike Information Criterion and BIC is the Bayesian Information Criterion. DM#2 is the best fitting model that was presented in the main text.

Experiment 1

Varied Parameters				Fixed Parameters	
Test Block	v_{WT}	v_{ST}	a	v_{WF}	-0.180
1	0.170	0.295	0.182	v_{SF}	-0.270
2	0.159	0.234	0.183	s_z	0.244
3	0.001	0.196	0.179	T_{er}	0.588
4	0.001	0.165	0.174	s_T	0.285
5	-0.078	0.132	0.173	η	0.155

Experiment 2

Varied Parameters				Fixed Parameters	
Test Block	v_{WT}	v_{ST}	a	v_{WF}	-0.235
1	0.118	0.351	0.175	v_{SF}	-0.252
2	0.092	0.315	0.181	s_z	0.155
3	0.037	0.261	0.181	T_{er}	0.579
4	0.003	0.176	0.177	s_T	0.291
5	-0.062	0.166	0.173	η	0.206

Experiment 3

Varied Parameters				Fixed Parameters	
Test Block	v_{WT}	v_{ST}	a	v_{WF}	-0.159
1	0.124	0.255	0.167	v_{SF}	-0.208
2	0.090	0.258	0.167	s_z	0.134
3	0.027	0.200	0.163	T_{er}	0.541
4	0.030	0.173	0.160	s_T	0.252
5	0.007	0.139	0.160	η	0.242

Table B.8: The diffusion model parameters of the starting point model for Experiment 3.

Appendix C

Analyses of the Encoding Task at Study

In this project, the levels of processing effect was used to manipulate strength at encoding. During study, participants were given either a shallow or a deep encoding task and these tasks were manipulated either across or within study lists. In the shallow task, participants were asked to make an orthographic judgment and for the deep task they were asked to make a semantic judgment about the study item. In this section, the association between the levels-of-processing task and the reaction time at study was investigated. Further, the associations between responses at study and performance at test was examined.

In a seminal paper, Craik and Tulving (1975) showed that the reaction times were longer when participants made a semantic judgment compared to the reaction times from an orthographic judgement. They suggested that the increase in the reaction time for the semantic judgments provided evidence for deeper level of processing as participants spent more time to perform the task. Further, they investigated the association between the type of the response and the reaction times. Their results showed that the reaction times were comparable between the negative ('no') and positive ('yes') responses and did not reveal a significant effect. In the current project, we investigated whether the type of the response

and the type of the encoding task was associated with the reaction time of those responses. We conducted a 2 (Encoding Task: Shallow, Deep) $\times$ 2 (Response: Yes, No) repeated measures ANOVA on the average reaction time data of individuals from the study session of Experiments 1-3. The main effect of the encoding task was significant in all three experiments, $F(1, 27) = 87.21$, $p < .001$; $F(1, 25) = 50.07$, $p < .001$; and $F(1, 21) = 78.34$, $p < .001$, respectively. As Figure C.1 shows the average reaction time was slower when participants were given a deep encoding task and this finding was consistent with the results from Craik and Tulving (1975). The main effect of response was significant only in Experiment 1, $F(1, 27) = 87.21$, $p = .038$, showing that the negative responses ($M = 0.94$, $SD = 0.26$) were significantly slower than the positive responses ($M = 1.01$, $SD = 0.34$), $t(55) = -2.39$, $p = .02$. In none of the experiments the interaction between the encoding task and the response reached significance.

In addition to reaction time of responses at study, we investigated the association between the response at study and recognition performance. Craik and Tulving (1975) analyzed the effect of encoding task on the hit rates conditional on the initial response given at study. In their results, items that had received a positive response were recognized better than the items that had received a negative response at study. This response effect was greater in magnitude as the encoding task was deeper. They argued that positive responses facilitates elaborative encoding. Coupled with the null effect of the response on the reaction times at study, the response effect provides evidence for the qualitative differences at encoding rather than the processing time. The results from the current project further supports this argument. A 2 (Encoding Task) $\times$ 2 (Response) repeated measures ANOVA was conducted on the hit rates at test. The main effect of encoding task was significant in all three experiments, $F(1, 27) = 258$, $p < .001$; $F(1, 25) = 109.4$, $p < .001$; and $F(1, 21) = 72.38$, $p < .001$, respectively. Consistent with the strength effect from the accuracy results, items that had received the deep encoding task produced greater hit rates compared to the items that had received the shallow encoding task. The main

effect of response was also significant across all three experiments, $F(1, 27) = 20.41$, $p < .001$; $F(1, 25) = 37.24$, $p < .001$; and $F(1, 21) = 14.836$, $p < .01$, respectively. Items that have received a positive response at the initial study task had greater hit rates than the items that had received a negative response (Figure C.2). These findings were consistent with the findings from Craik and Tulving (1975) paper. The interaction between response and the encoding task did not reach significance in any of the experiments.

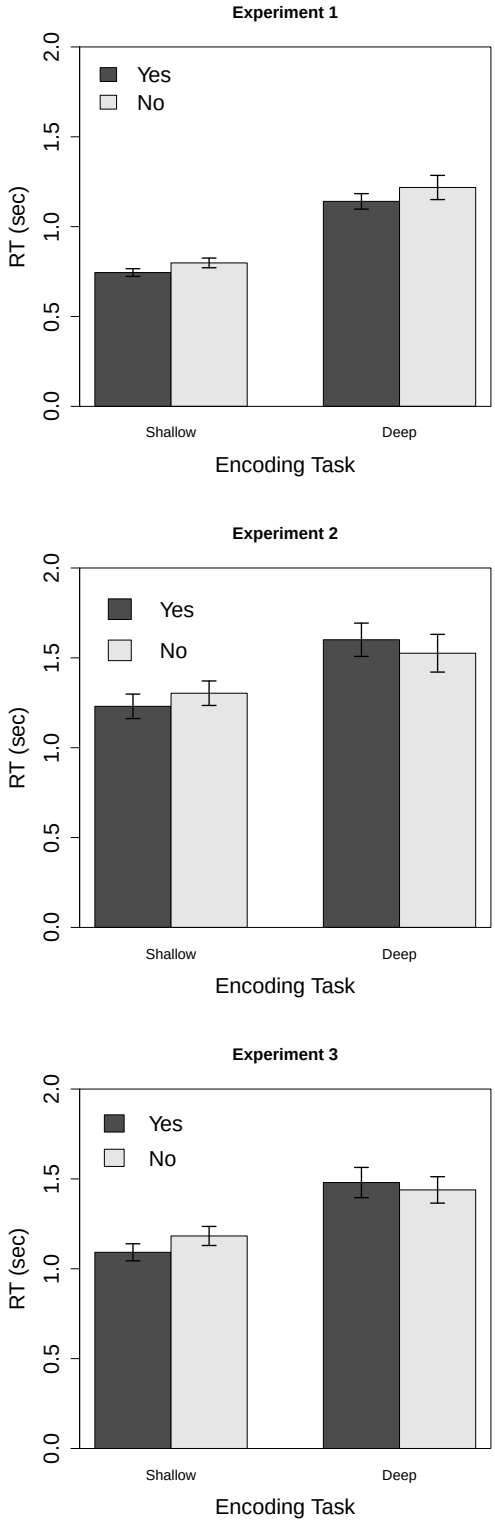

Figure C.1: Reaction time at study as a function of encoding task and response

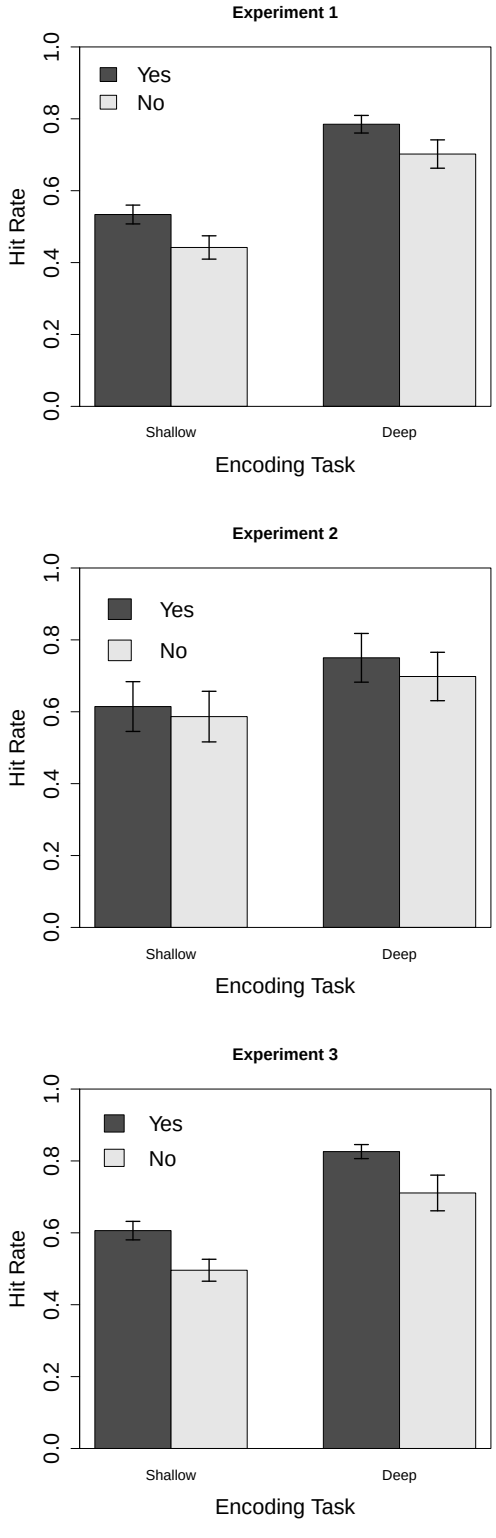

Figure C.2: Reaction time at study as a function of encoding task and response

References

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*(5), 681–692. doi: 10.3758/s13421-012-0185-5

Bates, D., Maechler, M., & Bolker, B. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39. [Computer software manual]. Retrieved from `http://cran.r-project.org/package=lme4`

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

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. doi: 10.1177/0049124104268644

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, *49*(2), 231–248. doi: 10.1016/S0749-596X(03)00061-5

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33A*(4), 497–505. doi: 10.1080/14640748108400805

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*(1), 135–150. doi: 10.1111/j.1756-8765.2011.01177.x

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory

research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. doi: 10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. doi: 10.1037/0096-3445.104.3.268

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. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology*, *59*(4), 297–319. doi: 10.1016/j.cogpsych.2009.07.003

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

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*(4), 447–460. doi: 10.1016/j.jml.2006.06.003

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478. doi: 10.1037/0033-295X.108.2.452

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 414–435. doi: 10.1037/0278-7393.27.2.414

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*(3), 145–154. doi: 10.1037/h0048509

Gibson, E. J. (1940). A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, *47*(3), 196–229. doi: 10.1037/h0060582

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32–41. doi: 10.1037/h0048826

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review; Psychological Review*, *91*(1), 1–67. doi: 10.1037/0033-295X.91.1.1

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

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

Hintzman, D., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition*, *8*(4), 378–382. doi: 10.3758/BF03198278

Hintzman, D. L. (1984). Minerva 2: A simulation model of human memory. *Behavior Research Methods*, *16*(2), 96–101. doi: 10.3758/BF03202365

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

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. doi: 10.1006/jmps.2001.1388

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*(5), 852–857. doi: 10.3758/BF03196776

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*, *52*(4), 493–504. doi: 10.1016/j.jml.2005.01.007

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. doi: 10.1126/science.1152408

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 773–795. doi: 10.1080/01621459.1995.10476572

Kiliç, A., Criss, A., & Howard, M. (2012). A causal contiguity effect that persists across time scales. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0028463

Kucera, F., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi: 10.3758/BF03210951

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*(4), 335–384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General*, *14*(4), 233–259. doi: 10.1037/a0025277

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

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–336. doi: 10.1037/0278-7393.31.2.322

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

Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434-455. doi: 10.1037/0033-295X.95.4.434

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *86*(6), 609–626.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review*, *90*(4), 316–338. doi: 10.1037/0033-295X.90.4.316

Murdock, B. B. (1993). Todam2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*(2), 183–203. doi: 10.1037/0033-295X.100.2.183

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The loyola symposium* (p. 145-194). Hillsdale, New Jersey: Erlbaum.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415. doi: 10.1037/0096-3445.128.4.403

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review; Psychological Review*, *88*(2), 93-134. doi: 10.1037/0033-295X.88.2.93

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59. doi:

10.1037/0033-295X.85.2.59

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin*, *86*(3), 446–461. doi: 10.1037/0033-2909.86.3.446

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, *92*(2), 212–225. doi: 10.1037/0033-295X.92.2.212

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., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review; Psychological Review*, *83*(3), 190. doi: 10.1037/0033-295X.83.3.190

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424. doi: 10.1016/j.jml.2003.11.002

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Rulletin & Review*, *9*(3), 438–481. doi: 10.3758/BF03196302

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261-300. doi: 10.1037/0033-295X.106.2.261

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 91–105. doi: 10.1037/0278-7393.6.1.91

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: Ii. theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 179–195. doi: 10.1037/0278-7393.16.2.179

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

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0028151

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. (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

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Tulvig, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology*, *92*(3), 297–304. doi: 10.1037/h0032367

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E., & Arbuckle, T. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 321–334. doi: 10.1016/S0022-5371(63)80012-2

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 145–150. doi: 10.1037/h0023344

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

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*(6), 1011–1026. doi: 10.3758/BF03193087

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*(1), 61–72. doi: 10.3758/BRM.40.1.61

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

Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Behavioral Monographs*, *1*, 1-82.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. doi: 10.3758/BF03196615

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. doi: 10.1016/j.jml.2007.04.006

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4),

442. doi: 10.1037/0278-7393.1.4.442

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory1. *Journal of Verbal Learning and Verbal Behavior*, *2*(5-6), 440–445. doi: 10.1016/S0022-5371(63)80045-6

VITA

NAME OF AUTHOR: Aslı Kılıç

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Bilkent University, Ankara, Turkey

Middle East Technical University, Ankara, Turkey

Syracuse University, Syracuse, NY

DEGREES AWARDED:

Doctorate of Experimental Psychology, 2012, Syracuse University

Master of Science in Cognitive Sciences, 2007, Middle East Technical University

Bachelors of Science in Management, 2004, Bilkent University

AWARDS:

Syracuse University Eric F. Gardner Fellowship recipient, 2011-2012

Syracuse University Travel Grants, Fall 2009, Spring 2010, Fall 2011

PROFESSIONAL EXPERIENCE:

Teaching Assistant, Department of Psychology, Syracuse University 2007-2012

Research Assistant, Department of Psychology, Syracuse University 2007-2012

Research Coordinator, Department of Psychology, Syracuse University 2008-2009

Teaching Assistant, Department of Cognitive Sciences, Middle East Technical Univeristy 2006-2007