January 2017

Effects of Item Relatedness on Output Interference in Recognition Memory

Jennifer Sloane
Syracuse University

Abstract

In recognition memory tasks, output interference (OI) effects manifest as a decrease in performance over the course of a test list. While this interference effect has been shown in many different experimental contexts, it is unclear how it is influenced by different properties of study and test items. This work investigates the relationship between semantic similarity and OI in memory, comparing semantically related (animal names or emotion words) and unrelated items to better understand memory and decision processes. Because the related items share similar features and are more confusable, it was predicted that there would be a greater amount of OI for related items. The first experiment used a single item recognition task and showed OI for each stimulus type, but no difference between related and unrelated words. However, there were differences in memory bias amongst the emotion condition that make it difficult to interpret discriminability measures. To remove the potential confound of memory bias, Experiments 2 and 3 use a two-alternative forced choice version of the task. There were mixed findings for the effects of OI in these experiments, such that there were no significant effects in the emotion condition and weak to moderate effects in the animal condition. Additionally, related words only produced a greater amount of OI in the final experiment and only in the animal condition. These results show that related features for emotion words lead to an increased memory bias but no difference in OI, whereas related features for animal names lead to an increase in OI but no difference in memory bias. These results are discussed with potential implications to current theories and models of recognition memory processing.

EFFECTS OF ITEM RELATEDNESS ON OUTPUT INTERFERENCE IN RECOGNITION MEMORY

by

Jennifer Sloane

B.A., University of Maryland College Park, 2014

Thesis
Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Experimental Psychology.

Syracuse University
June 2017

Table of Contents

List of Tables

List of Figures

Introduction

Memory is an important part of everyday life that has complex effects on behavior. Recognition memory tasks are commonly used as a method to test memory and probe the processes involved in memorial decisions. The current studies seek to improve our understanding of memory and decision processes by specifically focusing on two factors that robustly affect recognition memory performance: item relatedness, which typically produces as a bias to believe related items were studied (Dougal & Rotello, 2007; White, Kapucu, Bruno, Rotello, & Ratcliff, 2014), and OI, which typically produces as a decrease in performance over the course of a test list (Criss, Malmberg, & Shiffrin, 2011; Schulman, 1974). It should be noted that although the finding of a memory bias for related items is especially robust for emotion stimuli (Dougal & Rotello, 2007; Kensinger & Corkin, 2003; Sharot, Delgado, & Phelps, 2004), it will also be investigated using animal names in these experiments. The goal was to determine whether properties of the studied/tested items (i.e., their relatedness) influenced the rate of OI at test. The following sections describe first, how memory performance is measured in recognition tasks, and second, how these measures relate to effects of item relatedness and OI. Then a series of experiments are presented that assess whether and how these two factors interact. The paper concludes with discussion of how these results impact current theories of recognition memory processing.

Discriminability and Bias in Memory Tasks

In single-item recognition tasks, studied and unstudied items are presented and participants must decide whether each one is “old” (studied) or “new” (unstudied). Data from these tasks are typically presented in terms of hits, or the proportion of studied items that were correctly labeled old, and false alarms, or the proportion of unstudied items that were incorrectly labeled old (Figure 1a). From these data, performance can be assessed with two primary measures: bias and discriminability. An old bias refers to the overall likelihood of calling items old at test, whereas discriminability refers to the ability to successfully discriminate old from new items. Measures of these components can be estimated from data using the signal detection theory framework (Green & Swets, 1966; Wixted, 2007). There are different measures and assumptions that could have been used to calculate bias and discriminability, but the signal detection theory framework is a common approach and is what will be used here. The most common measure for discriminability is d', which measures the distance between the means of the two distributions and is expressed by:

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

Where *z* indicates normalized values for the hit rates (HR) and false alarm rates (FAR). A high d' is indicative of being able to accurately discriminate between studied items and unstudied items and thus reflects strong memory performance (Figure 1b).

Also taken from the signal detection theory framework is the measure of bias, which is a way to express an individual's decision criterion and computed by the following equation:

$$C = - \frac{z(HR) + z(FAR)}{2}$$

A C of zero is indicative of no bias to call any given word either old or new. From this equation, a positive bias means a conservative criterion that increases the likelihood of calling a word new and a negative bias means a liberal criterion that increases the likelihood of calling a word old (Figure 1c).

The use of this signal detection framework has the advantage of dissociating effects of bias and discriminability. However, one limitation of analysis with d' and C is that these two measures are not completely independent. Specifically, differences in bias can make it appear as though there are differences in discriminability even when there are none (Dougal & Rotello, 2007). One possible way to circumvent this problem is by using receiver operating characteristics (ROC) curves to independently observe discriminability and bias (Macmillan & Creelman, 1991). However, in order to calculate ROC curves, confidence rating must also be collected after every test item. Because the main interest of this project was to observe effects of OI on related and unrelated items, the data were already divided into different bins and separated by stimulus type. Adding confidence ratings would result in further dividing the data, which would mean even fewer data points within each confidence rating. Another way to avoid this problem is to use two-alternative forced choice (2AFC) recognition tasks, wherein two items are presented and the participant must decide which one is old. In this task, the bias is eliminated because the overall likelihood of considering items old does not affect the decision (since there is always one old item). In 2AFC tasks, discriminability is taken as overall accuracy and there is no measure for bias. Because the 2AFC task does not require any further subdivision of the data, this was the method that was chosen.

With these measures of performance in mind, the next section focuses on how the factors of item relatedness and OI affect memory performance. As described below, relatedness typically manifests in measures of bias, whereas OI manifests in measures of discriminability.

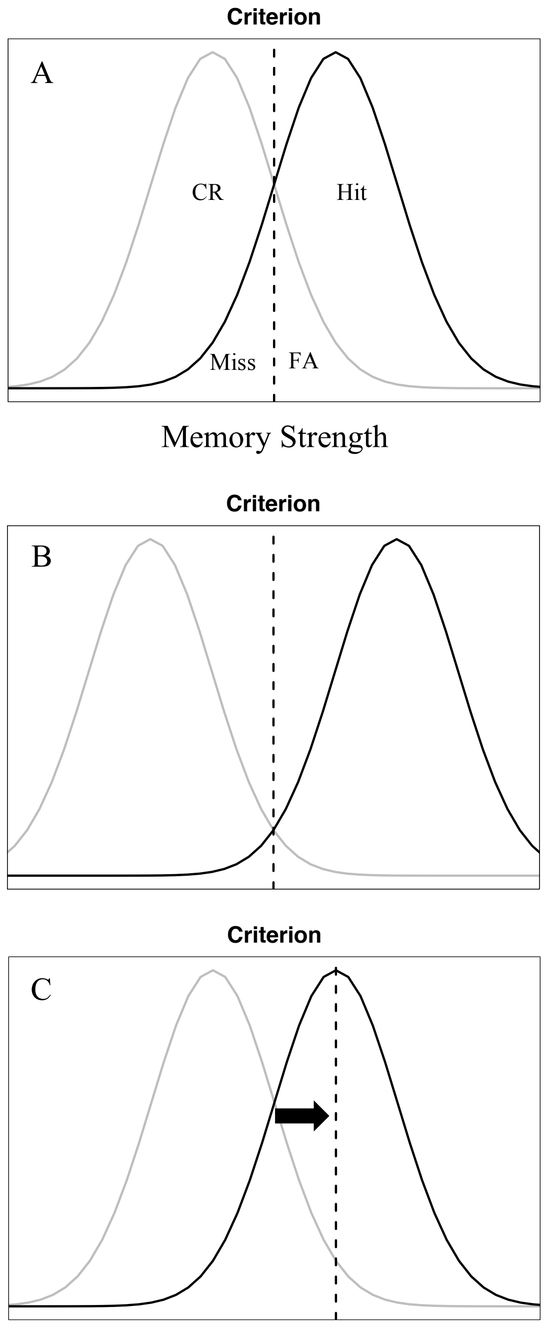

Figure 1. a) Signal Detection Theory (SDT) model depicting hits, false alarms (FA), misses, and correct rejections (CR). The grey represents the new distribution, the black represents the old distribution, and the dotted line is the decision criterion. b) A SDT model illustrating a higher d' depicted by less overlap of the two distributions. c) A SDT model illustrating a more conservative decision criterion.

Effects of Item Properties on Memory Performance

Different types of items or stimuli affect how we make memory-based decisions. There are a myriad of item properties that have been shown to influence memory performance (e.g., word frequency effects (Monaco, Abbott, & Kahana, 2007)), but this study focuses on the robust finding that items that are related to each other tend to affect memory differently than those that are unrelated (Goh & Goh, 2006; Roediger & McDermott, 2004; Shiffrin, Huber, & Marinelli, 1995; White et al., 2014). Studying and testing items that are semantically related leads to a greater number of false alarms compared to studying items that are unrelated (Montefinese, Zannino, & Ambrosini, 2015). The Deese-Roediger-McDermott paradigm involves presenting participants with a list of semantically related words (i.e. dream, bed, pillow), but excluding a critical item of the category (i.e. sleep). There are robust findings demonstrating that a majority of participants will falsely recognize the critical item as old at test (Nabeta & Kusumi, 2009; Roediger & McDermott, 2004). This suggests that semantically related items can negatively affect memory performance and can often lead to false memories or false alarms because the relatedness of the items make them difficult to discriminate at test. However, related items also lead to higher hit rates compared to unrelated items (White et al., 2014), thus the effect of item relatedness is usually seen in a stronger bias to label those items old.

Item information can also affect memory decisions through defining characteristics of the items themselves. There is an especially large literature on emotion stimuli. Specifically, emotional valence, arousal, and dominance have been shown to influence memory decisions. Although the main focus of this paper is on related words in general, and not specifically emotion, what follows is a brief overview of some of the findings within the emotion literature. There is overwhelming support for a memory bias for emotion items (Dougal & Rotello, 2007).

In other words, participants are likely to recognize more emotion items as old, whether or not they studied the words originally. For example, if the study list included negatively arousing words, such as “trauma, bed, blood, tree, fear”, participants would correctly identify trauma, blood, and fear as old, but also incorrectly identify unstudied emotion words (e.g. anxiety, danger). While there is extensive literature probing this phenomenon (Adelman & Estes, 2013; Doerksen & Shimamura, 2001; Kensinger & Corkin, 2003), the specific factors contributing to this bias are still being debated (Gallo, 2010).

That being said, it is clear that this memory bias is a result of some aspect of the semantic similarity of these words. First, several studies have found an emotional memory enhancement effect, such that participants are more likely to recall or recognize emotional stimuli compared to neutral stimuli (Buchanan & Adolphs, 2002; Hamann, 2001; Kensinger & Corkin, 2003). Second, the emotion words are part of a unique category and share similar features. White et al. (2014) tested this by using a non-emotion category of animal names. Findings showed a memory bias for animal names in the same direction as emotion names, but the effect was not quite as strong. This suggests that although the bias to call negative emotion words old is driven by effects of categorical membership (White et al., 2014), there may still be a unique quality about emotion items that make them even more susceptible to this memory bias. As there was a weaker effect for the memory bias in animal names, the first experiment presented serves as a replication of White et al. (2014). Additionally, the effects of categorical membership were further explored in all three experiments by including both negative emotion words and animal names, using the same word pool as White et al. (2014).

Although there are reliable findings for a memory bias for emotion stimuli, there are inconsistencies in the literature for the effects of discriminability. While some studies find enhanced d’ for emotion items (Kensinger & Corkin, 2003), others found no difference (Doerksen & Shimamura, 2001), and still others found a decrease in d’ (Dougal & Rotello, 2007). This has led to an ongoing debate in the literature revolving around whether and how emotion influences discriminability (White et al., 2014).

It is clear from the literature that item properties like relatedness and valence affect memory performance (Anders, Lotze, Erb, Grodd, & Birbaumer, 2004; Kensinger & Corkin, 2003). The effects consistently manifest in measures of bias, whereas effects on measures of discriminability are inconsistent. This work further explores the measures of bias and discriminability within emotion stimuli as well as animal names. However, as described below, discriminability is also influenced by testing effects of OI, and it is unknown whether item relatedness interacts with OI. If this interaction exists, it could potentially influence bias and discriminability in unaccounted for ways.

Output Interference Effects in Memory

In addition to effects related to item characteristics like similarity and valence, memory decisions can be influenced by factors related to the memory test itself. Although there are countless studies on testing effects and the benefits to testing oneself (Glover, 1989; Hogan & Kintsch, 1971; Roediger & Karpicke, 2006), there are also disadvantages to testing, especially if it is over an extended period of time or long test list. There is robust evidence for OI, which is the finding that memory performance, as measured by discriminability, decreases over the course of a test list (Aue, Criss, & Prince, 2015; Malmberg, Criss, Gangwani, & Shiffrin, 2012; Norman & Waugh, 1968). As items are encoded into memory they produce interference with other items already stored in memory, which ultimately decreases performance during test. Murdock &

Anderson (1975) showed that the number of items shown at test is negatively correlated to performance. Another one of the earliest studies on OI in recognition memory used a 2AFC task and similarly found that recognition fell by 55-75% between the first and last quarter of the test list (Schulman, 1974).

Studying OI can help to answer some of the discrepancies of discriminability from studies of item relatedness. If related items produce more (or less) OI than unrelated items, this could confound measures of discriminability that do not account for OI effects. Take for example a study that found equal discriminability for emotion (related) and unrelated words (e.g., Dougal & Rotello, 2007; White et al., 2014). If OI was different between these items, it could mask actual differences in discriminability. For example, it is possible that at the beginning of test, discriminability for related words was greater than unrelated, but there was a greater effect of OI on related items over the course of the test (see Figure 2). Thus, discriminability would be higher for related items at the beginning of the list, but lower at the end. The early advantage for related items would be canceled out by the later disadvantage. Traditional analyses include only the overall response rate and do not observe how these rates change over the course of the test list. This would make it impossible to see what is happening over time because the effects of OI could mask the enhanced discriminability for related items. Thus accounting for the effects of OI in these studies could reveal underlying differences that are not apparent in traditional analyses of recognition data.

Overall, it is important to simultaneously investigate item relatedness and OI and how they affect bias and discriminability to better understand the nature of memory and decision processes. This paper presents three recognition memory experiments which investigate whether and how item relatedness interacts with OI.

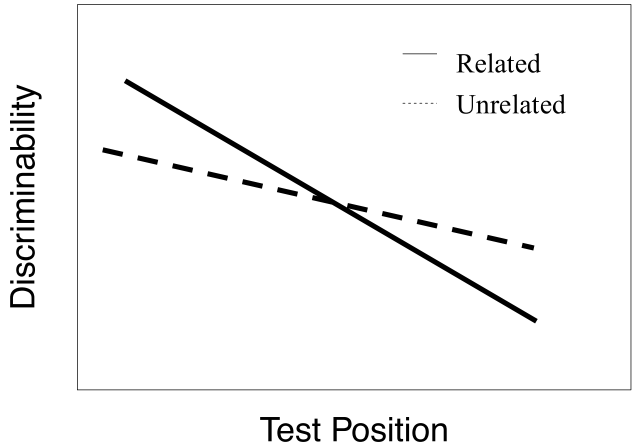

Figure 2. A hypothesized interaction showing that although related stimuli may start off with better discriminability compared to unrelated stimuli, over the course of the test list there is a greater effect of output interference on related stimuli resulting in lower discriminability at the end of the test for related items.

Experiments

These experiments seek to provide a deeper understanding of memory and decision processes by investigating the interplay among item relatedness and OI at test. All participants completed a recognition memory task that included a study list, a short math distractor task, and a test list. In the first experiment, participants completed a single item test phase, where they saw one item at a time (half old and half new words) and had to identify if the word was previously studied or novel. The second and third experiments consisted of a 2AFC test phase, where participants were presented with two items at test and had to identify the previously seen item. All experiments and conditions consisted of mixed study and test lists, such that half the items were unrelated and half the items were related (i.e. negative emotion or animal stimuli) and presented in a randomized order. Results were analyzed relative to differential effects of OI for related and unrelated items. The primary method of analysis was to divide the data into bins and observe the effects of discriminability across test position as stated in the preregistration (see appendix A). However, two follow up analyses were also completed for the force choice experiments using a trial by trial linear regression and mixed effect models.

Experiment 1

Procedure

The first experiment was a single item recognition memory task, where participants were presented with 96 study items one at a time. Each item was displayed on a computer screen for three seconds. For each study item, participants responded to whether or not the letter "E" was present. This encoding task was done to ensure participants were paying attention during the long study phase. All items were presented in a random order. Then, participants completed 15 short math distractor problems. At test, participants responded to 192 items, half targets and half lures, and indicated whether the word was old or new by pushing either the "z" or "/" keys. Items were presented one at a time at test and participants had a maximum of three seconds to respond before the next item. Participants were randomly assigned to one of two groups based on the type of words used in the task: a mixed emotion or mixed animal condition, meaning the related and unrelated stimuli were randomized. Only mixed conditions were used in these studies in order to see how related and unrelated items interact and effect OI throughout the entire test list. It would be interesting in future work to include pure related and pure unrelated conditions. However, that would address a slightly different question of observing only one type of stimulus, which is beyond the scope of this particular project.

Stimuli

Stimuli consisted of a matched set of negative-emotion and unrelated words or a matched set of animal names and unrelated words for all experiments. One pool contained 96 negative emotion words and 192 matched unrelated words, which were chosen from the ANEW pool and consistent with the stimuli used in Dougal and Rotello's (2007) study. On a nine-point scale, the emotion and unrelated stimuli differed in both arousal ($M_{emotion} = 6.63$, $M_{unrelated} = 4.15$) and valence ($M_{emotion} = 2.24$, $M_{unrelated} = 5.16$). Word frequency (Francis & Kucera, 1982) and semantic interrelatedness using latent semantic analysis (Landauer, Foltz, & Laham, 1998) were matched across the word pools. The second pool contained 96 animal names and 189 matched unrelated (excluding animal names in the unrelated list) words, taken from the Van Overschelde, Rawson, and Dulosky (2004) database and consistent with the stimuli in White et al. (2014 Exp. 2). These two word pools did not differ in arousal [$M_{animal} = 4.95$, $M_{unrelated} = 4.89$, $t(15) = .73$, $p = .48$] or valence [$M_{animal} = 5.13$, $M_{unrelated} = 5.01$, $t(15) = 1.3$, $p = .21$] (White et al., 2014). These analyses of the stimuli are simply to show that the emotion stimuli contain unique features (i.e. valence and arousal) that may separate them even more from other categories (i.e. animal names). However, it should be noted that the focus is not on the emotional aspects of the words, but rather whether the stimuli are related and part of a category or unrelated. In all three experiments, words were randomly selected to be used in the two conditions. The two word pools were used to observe effects of OI in different categories.

Participants

Syracuse University undergraduates participated in this experiment for course credit. The total sample size was 43 participants for the emotion condition and 39 for the animal condition.

However, 6 participants from the emotion condition and 10 from the animal condition were excluded for either having a d’ measure (HR - FAR) less than .10 or for not responding to over half of the test items, leaving 37 and 29 participants, respectively, for analyses. Additionally, individual trials were excluded if the RT was shorter than 250 ms; these cutoffs were consistent throughout all the experiments.

Hypotheses

Experiment 1 looked at effects of item relatedness on both bias and discriminability for related and unrelated words with two specific hypotheses. It was first predicted that there would be a memory bias for related items as this is consistent within the literature (Dougal & Rotello, 2007; White et al., 2014). As more and more semantically similar items are presented at test, there is a greater chance for interference and for the items to become confusable compared to unrelated stimuli. Participants would, then, be more likely to call related words old and would have a harder time discriminating between studied and unstudied words with the progression of the test. Therefore, it was also predicted that there would be a greater amount of OI for related items compared to unrelated items over the course of the test.

Results

It is typical to first observe the descriptive statistics, including the overall hit and false alarm rates. A summary of the response rates is shown in Table 1. The related words in the emotion condition had significantly more hits, $t(35) = 9.13$, $p < 0.001$, and false alarms, $t(35) = 7.74$, $p < 0.001$, than the unrelated words, suggesting a bias to call negative emotion words old. However, when comparing the related and unrelated words in the animal condition, there was neither significantly more hits, $t(27) = 0.66$, $p = 0.510$ nor false alarms, $t(27) = -1.01$, $p = 0.312$

(Figure 3). It should be noted that all error bars were computed by taking the 95% confidence interval. Because of the opposing patterns of behavior seen here, the data was kept separated by conditions for the remaining analyses.

Since the focus of the study is on testing effects, the data were divided into four bins of 48 trials each, in order to see how discriminability and bias change over the course of the test list (Figure 5). Two 4 (bins) x 2 (stimulus type) repeated measures ANOVAs were performed on both the d' and bias data, separately. First looking at d', there was a main effect of bin number for emotion, $F(3,105) = 9.41$, $p < .001$ and animal words, $F(3, 84) = 4.85$, $p = 0.004$, showing the effect of OI across the test list. However, there was no main effect of stimulus type (i.e. related and unrelated), suggesting discriminability was not differentially affected by related and unrelated words. There was also no significant interaction between bin number and stimulus type, meaning OI was not greater for related words compared to unrelated words. Looking at bias next, there was again a main effect on bin number for emotion, $F(3,105) = 4.57$, $p = 0.012$ and animal names, $F(3, 84) = 11.16$, $p < .001$. The plots in Figure 5 illustrate this main effect was manifested as an increase in bias over the course of the test. There was also a main effect on stimulus type, but only for the emotion condition, $F(3,105) = 66.96$, $p < .001$. There was not a significant interaction for either emotion or animal stimuli, meaning there were no interactions between stimulus type and bin number.

The following statistical analyses were done using a paired sample t-test to compare the slopes of related and unrelated words using the discriminability and bias data. Results from the discriminability data were insignificant for both the emotion, $t(35) = -0.63$, $p = 0.538$ and the animal condition, $t(27) = 0.07$, $p = 0.941$, meaning there was not a stronger effect of OI on related words. Similarly, there were no significant differences when looking at the bias data for emotion, $t(35) = -0.82$, $p = 0.9335$ or animal stimuli, $t(28) = -0.13$, $p = 0.896$. Although bias appears to get increasingly more conservative over the course of the test list, these results indicate that there was no difference in the rate of change of bias when comparing the type of stimuli.

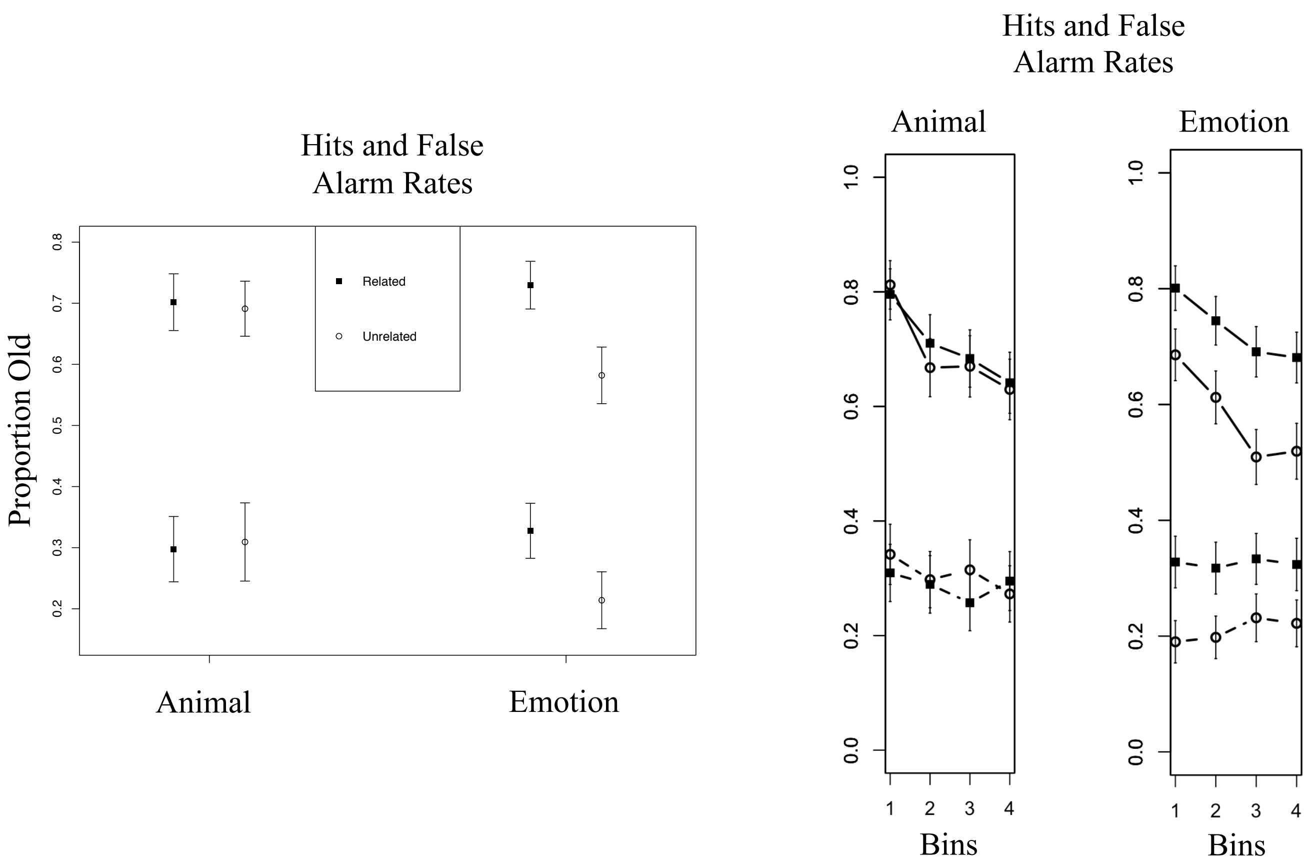

Figure 3. Hit and false alarm rates averaged across participants in both conditions and across the four bins of 48 trials each. On the right, hits are represented by solid lines and false alarms are represented by dashed lines.

Stimuli	Hits	FA	d’	C
Animal Pool				
Related	0.702 (0.12)	0.298 (0.14)	1.158 (0.52)	0.013 (0.34)
Unrelated	0.691 (0.12)	0.309 (0.17)	1.098 (0.57)	0.010 (0.36)
Emotion Pool				
Related	0.730 (0.12)	0.328 (0.14)	1.131 (0.43)	-0.087 (0.32)
Unrelated	0.582 (0.14)	0.214 (0.14)	1.092 (0.55)	0.329 (0.33)

Table 1. Summary of statistics for Experiment 1 where FA is the false alarm rate, d’ is the measure of discriminability, and C is the measure used for calculating bias for animal related, animal unrelated, emotion related, and emotion unrelated stimuli. Standard deviations are provided in parentheses.

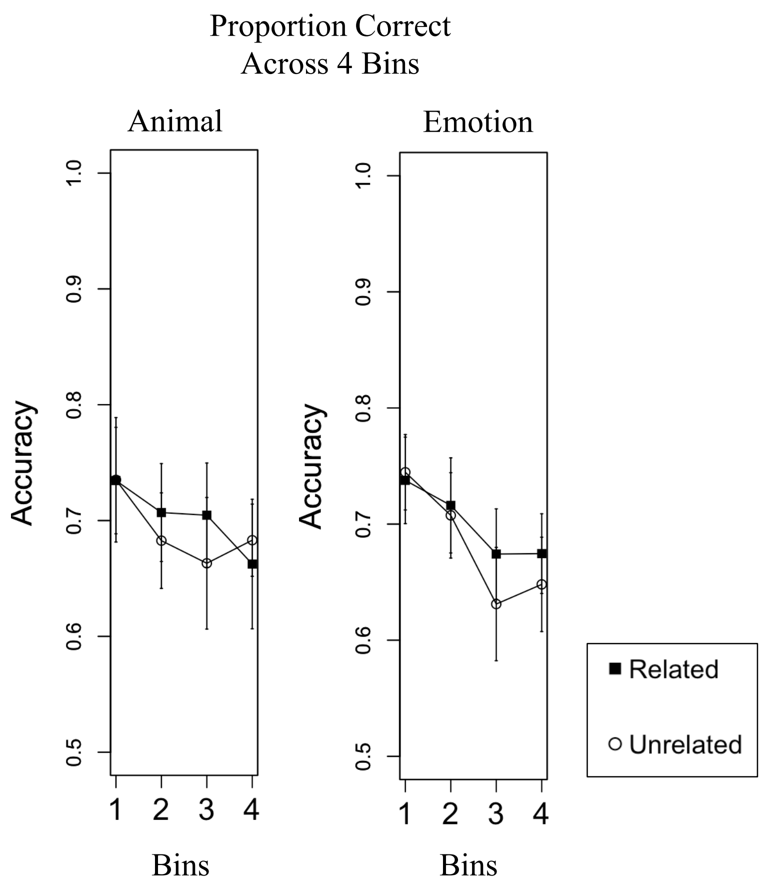

Figure 4. Average proportion correct for all related words and unrelated across four test bins of 48 trials each.

Stimuli	Proportion Correct: Bin 1	Proportion Correct: Bin 2	Proportion Correct: Bin 3	Proportion Correct: Bin 4
Animal Pool				
Related	0.735 (0.12)	0.707 (0.11)	0.705 (0.12)	0.662 (0.15)
Unrelated	0.735 (0.14)	0.683 (0.11)	0.663 (0.15)	0.683 (0.08)
Emotion Pool				
Related	0.736 (0.11)	0.714 (0.12)	0.679 (0.12)	0.675 (0.10)
Unrelated	0.750 (0.10)	0.707 (0.11)	0.631 (.15)	0.648 (.12)

Table 2. Mean proportion correct responses for related and unrelated stimuli per bin for both animal and emotion stimuli with standard deviations reported in parentheses.

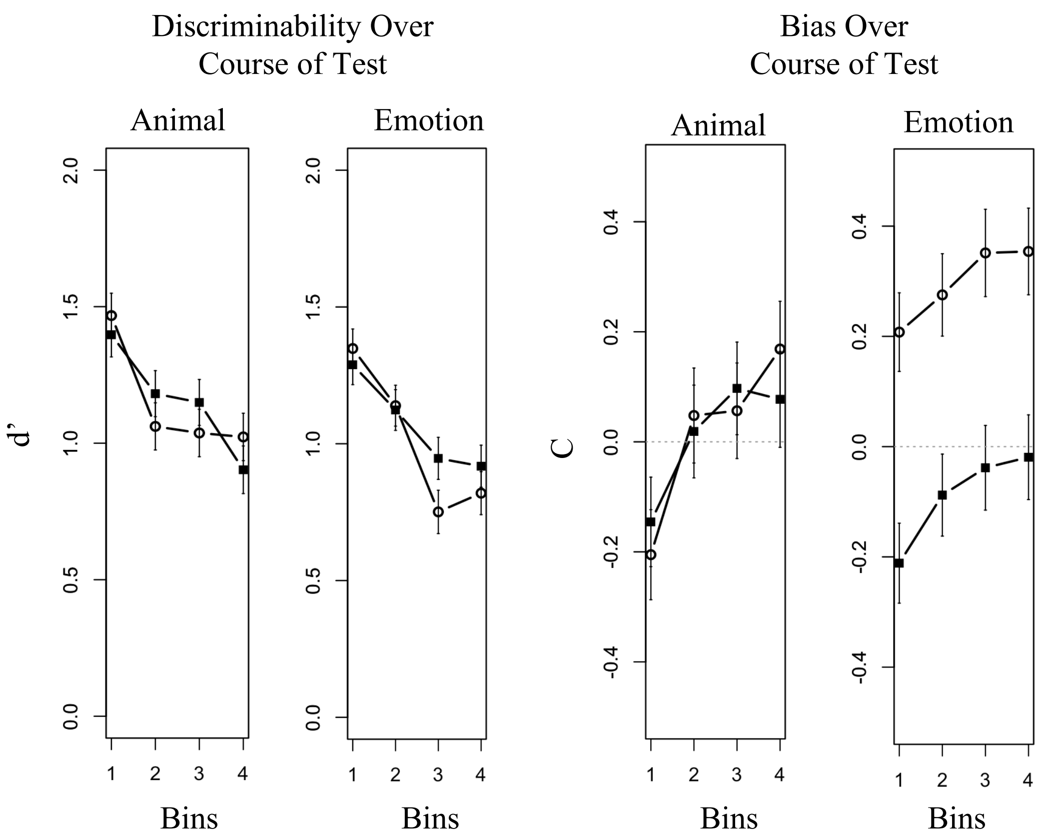

Figure 5. Signal detection theory measures of d' (top panel) and bias (bottom panel) divided into four bins of 48 trials to see change over the course of the test list.

Discussion

Experiment 1 explored if OI is differentially affected not only by different types of stimuli (related vs. unrelated), but also by different types of categories (emotion vs. animal). When studying testing effects in single item recognition memory paradigms, it is crucial to observe performance in terms of discriminability and bias over the course of a test list; this was done by dividing the test trials into four bins. Results from Experiment 1 provide further evidence for the significant effects of OI, since discriminability decreased across the test list in both conditions and both stimuli types. Interestingly, there was no support for the hypothesis of a greater increase in OI for related items over the course of the test list.

Additionally, results from this experiment found evidence supporting the hypothesis for a memory bias to call related items old, as there were significantly more hits and false alarms, but only for the emotion condition. Although White et al (2014) found that categorical membership may play a large role in this liberal bias for emotion items, this memory bias is not present in the animal condition.

The levels of processing task ("is there an E in this word") was included to try to ensure participants were staying focused throughout the study phase. However, it is possible that the lack of memory bias for animal names from the present study is due to the shallow encoding task that prevented encoding of the relevant categorical features; whereas emotion items possess other unique features that the bias is still present even with this shallower level of processing. In the following experiment, this is addressed by adjusting the study phase to include a deeper level of processing task. It should be noted that previous literature has shown that manipulating levels of processing does not negate the effects of OI (Kilic, 2012). Although there is evidence that a greater depth of processing leads to enhanced memory performance and a slower rate of decay for the memory trace compared to shallower levels of encoding (Craik & Lockhart, 1972), Kilic (2012) illustrates that both deep and shallow levels of encoding are subject to effects of OI.

However, the measure of discriminability used here, d', assumes the conditions have an equal amount of bias (Dougal & Rotello, 2007). This is problematic because d' can be contaminated by differences in bias, which is seen in the emotion condition where the related stimuli elicit a more liberal bias and unrelated words a more conservative bias. This bias, unfortunately, makes it difficult to determine whether OI effects are in fact similar for related and unrelated items. The following experiments utilize a 2AFC paradigm to circumvent this bias problem to get an even more accurate measure of discriminability.

Experiment 2

Participants completed a 2AFC recognition memory task designed to observe OI effects in mixed conditions to circumvent the bias problem of Experiment 1. The two conditions were exactly the same as Experiment 1, with half unrelated items and half related (either animal and negative emotion) stimuli. Aside from the slightly different task paradigm where two items are presented at test and the participant has to identify the old one, there is also a deeper level of encoding task during the study list.

Procedure

The study phase was identical to that of Experiment 1, except the levels of processing task was different. To ensure strong performance, the study list capitalized on the self-reference effect in memory, which is the phenomenon where memory is enhanced when information is encoding about one's self (Cunningham, Brebner, Quinn, & Turk, 2014). Therefore, for each study item, participants responded to the following question: "Does this item relate to you?" The levels of processing task was changed in this experiment to ensure the participants were encoding the item specific information, not just the structural features.

After a short distractor task, participants completed a test phase. In a typical 2AFC task, a pair of items is presented simultaneously on the screen and the participant is told to choose the old, or previously studied, item. However, there is new evidence suggesting participants may make absolute judgments on forced choice tasks. Starns et al. (Starns, Chen, & Staub, 2017) conducted an eye tracking study, showing that sometimes participants only pay attention to one word before making a decision, rather than looking at both words and making a relative decision. This could pose another bias problem. Therefore, the design for the 2AFC test phase is modified from the standard paradigm. Here, one word is presented at test for one second and removed from the screen, then the second word is presented in the same manner, and finally both words are presented together for up to three seconds. Participants can only respond once both words appear together. The next test pair begins as soon as a response is given or after the maximum time. Participants are tested on a total of 96 word pairs.

There are four different possible test pairs that could be presented assuming there is always one and only one correct answer: 1) related target and related foil 2) related target and unrelated foil 3) unrelated target and unrelated foil 4) unrelated target and related foil. Because the main interest in this paper is to explore if and how related and unrelated words are differentially affected by OI, only the two test pairs with consistent categories (i.e. related target and related foil; unrelated target and unrelated foil) were included. For example, one test trial may include two emotion stimuli (i.e. "death" and "blood"), whereas another trial may include two unrelated stimuli (i.e. "chair" and "book"). Using this design, there will always be one target and one foil item within the same category, so participants will have to rely on the old distribution to make their decisions. In the single item recognition task, participants were required to set a decision criterion of calling a word new or old because there was only one item presented at a time. However, in the forced choice paradigm, participants make a relative judgment, meaning they no longer need to set a criterion. Thus, memory bias is no longer an issue because a participant's overall tendency to endorse items as studied is no longer relevant to the decision and participants are forced to look at both items that are presented.

Participants

Syracuse University undergraduates participated in this experiment for course credit. Participants were randomly assigned to one of the two conditions: emotion or animal. Data were collected for 50 participants in each condition. However, 16 participants were excluded in the analyses for the emotion and 11 for animal condition due to not responding to a majority of the test questions or failing to complete the encoding task during the study phase. Leaving a total of 34 participants in the emotion and 39 participants in the animal condition.

Hypotheses

Similar to the hypothesis in Experiment 1, it was predicted that there would be effects of OI and there would be a greater amount of OI for related items compared to unrelated items.

Results

First, the data were divided into the different stimuli type, either related or unrelated, for both the animal and emotion condition. In 2AFC tasks, there are only correct and incorrect responses, unlike in single item recognition where there are hits, false alarms, misses, and correct rejections. Therefore, the discriminability measure is the proportion of correct responses. Similar to Experiment 1, the data were divided into four bins, 24 trials each, to observe if and how accuracy changes over the course of a long test. Because of the differences in bias between animal and emotion stimuli, as shown in Experiment 1, the two conditions will remain separate for all remaining analyses.

To observe how accuracy changed over the course of the test, a 2 (stimulus type) x 4 (bin number) repeated measures ANOVA was conducted on accuracy. Results found no main effect of bin number for emotion stimuli, $F(3,99) = 0.67$, $p = 0.571$ and a marginal effect for the animal stimuli, $F(3,114) = 2.707$ $p = .049$. These results suggest that there was no effect of OI in the emotion condition and only a change in OI from first to last bin in the animal condition. There was a main effect of stimulus type for emotion words, $F(1,33) = 5.16$, $p = .030$, but not animal names, $F(1,38) = 0.01$, $p = 0.939$. This shows that there was a difference between related and unrelated words, but only for the emotion condition.

Additionally, the mean proportion of correct responses per bin was calculated for every participant. Figure 6 plots the overall proportion correct per bin for related (red) and unrelated (black) words. The following analysis was done to observe if there were differential effects of OI on stimuli type. Using the average proportion of correct responses per bin for each individual, the slope for related and unrelated words were calculated and compared. A paired sample t test showed that there were no significant differences in emotion, $t(33) = -0.47$, $p = 0.644$, or animal stimuli, $t(38) = -0.83$, $p = 0.412$.

A linear regression model was used to observe performance on a trial by trial basis, rather than looking at the data grouped by bins. The individual data sets for related and unrelated words were used to calculate the average proportion correct responses for every test trial (i.e. 1-96) and the line of best fit was plotted (see Figure 7). In the animal condition, the correlation between proportion correct and trial number was -0.225 and -0.076 for related and unrelated words, respectively, and -0.174 and -0.171 for the emotion words. Although there were slight negative correlations for both related and unrelated words in both conditions, the only significant correlation was in the animal related stimuli, $t(94) = -2.23$, $p = 0.028$. This was similar to the results seen in the ANOVAs that there was only evidence of OI in the animal condition and specifically only in related stimuli.

However, in order to take random effects (i.e. individual differences) into account, an exploratory analysis was done using multi-level modeling. This was accomplished with lme4 package in R (Bates, 2015) to conduct a linear mixed-effects model and to explore the relationship between accuracy and test trial. The model comparisons discussed below were done for the animal and emotion conditions separately.

M_0, or the null model, had a fixed effect of stimulus type (i.e. related or unrelated) and random effect of stimulus type. To test the effects of OI, fixed and random effects of test position were added (M_1)[1]. The next model (M_2) observed the interaction between stimulus type and test position to see if OI affects related and unrelated words differently. Results of the model comparisons are shown in Table 4. Comparing M_0 to M_1 showed significant effects of OI in the animal, but not the emotion condition. However, comparing M_1 to M_2 showed that there were no differences in the amount of OI for related compared to unrelated words.

[1]One additional model was implemented using trial^2 as a fixed and random effect to account for nonlinearity. However, the results were consistent with the simpler model, so only the simpler model is presented.

Stimuli	Proportion Correct: Bin 1	Proportion Correct: Bin 2	Proportion Correct: Bin 3	Proportion Correct: Bin 4
Animal Pool				
Related	0.855 (0.15)	0.805 (0.21)	0.785 (0.20)	0.782 (0.25)
Unrelated	0.816 (0.13)	0.815 (0.22)	0.826 (0.16)	0.765 (0.22)
Emotion Pool				
Related	0.825 (0.20)	0.815 (0.18)	0.779 (0.20)	0.792 (0.19)
Unrelated	0.845 (0.22)	0.830 (0.20)	0.838 (0.16)	0.817 (0.17)

Table 3. Mean proportion correct responses for related and unrelated stimuli per bin with standard deviations reported in parentheses.

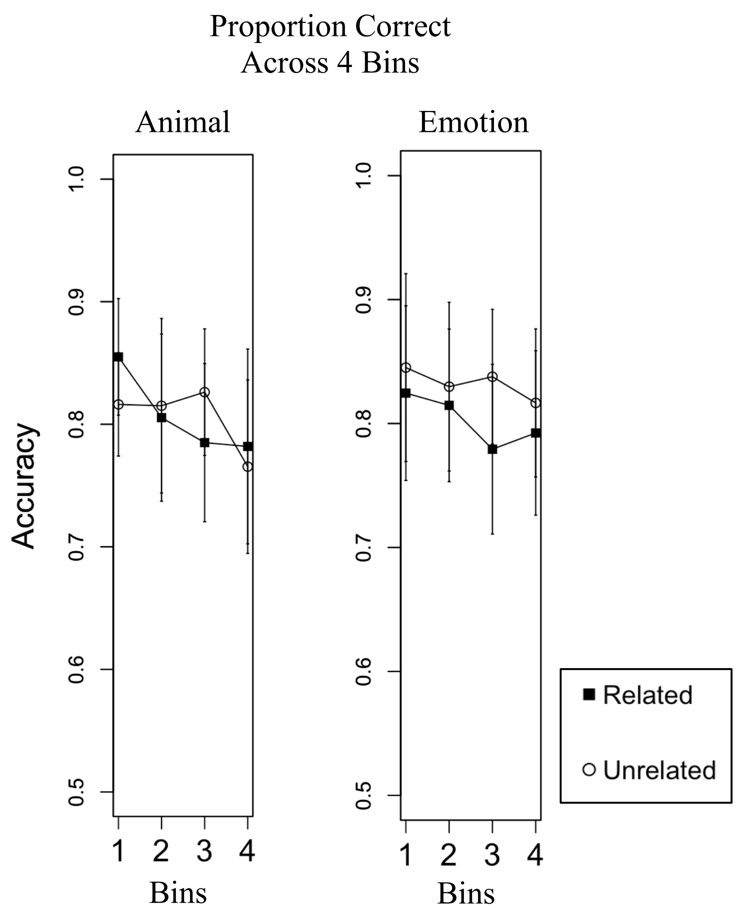

Figure 6. Average proportion correct for all related words and unrelated across four test bins of 24 trials each.

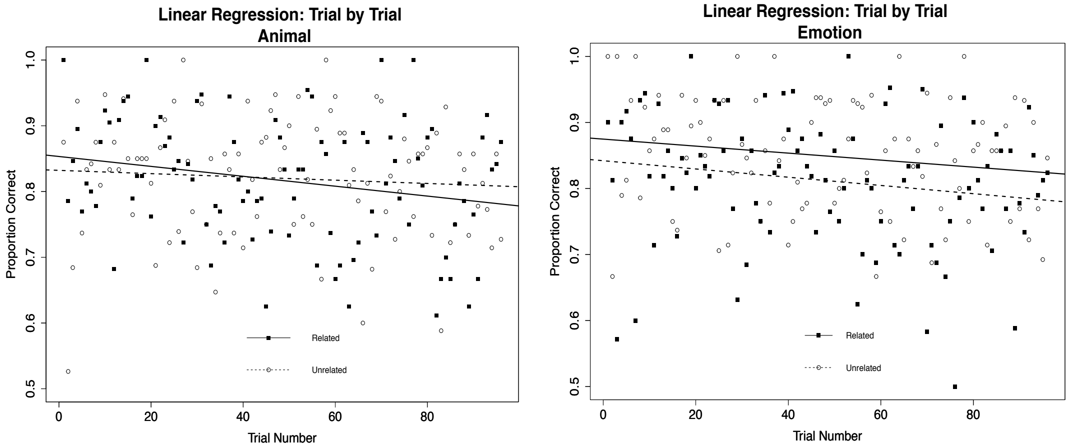

Figure 7. Trial by trial linear regression for proportion correct responses for all related word and unrelated, including lines of best fit.

Model Animal	**Fixed Effects**	**Random Effects**	**AIC**	**BIC**	**df**	χ^2
M_0	Stimulus Type	Stimulus Type	2888.1	2918.6	3267	
M_1	Stimulus Type, Test Position	Stimulus Type, Test Position	2877.3	2932.2	3263	M_1-M_0 = 18.82**
M_2	Stimulus Type, Test Position, Interaction	Stimulus Type, Test Position, Interaction	2878.9	2939.8	3262	M_2-M_1 = 0.42

Model Emotion	**Fixed Effects**	**Random Effects**	**AIC**	**BIC**	**df**	χ^2
M_0	Stimulus Type	Stimulus Type	2429.0	2459.0	2955	
M_1	Stimulus Type, Test Position	Stimulus Type, Test Position	2432.1	2486	2951	M_1-M_0 = 4.99
M_2	Stimulus Type, Test Position, Interaction	Stimulus Type, Test Position, Interaction	2433.8	2493.7	2950	M_2-M_1 = 0.25

*Table 4. The model comparisons for accuracy for the animal (top) and emotion (bottom) conditions. AIC is the Akaike Information Criterion and BIC is the Bayesian Information Criterion. The significant differences between χ^2 values of each model pair at α = 0.05 is represented by * and α = 0.001 is represented by ** sign.*

Discussion

There were two main goals of Experiment 2. The first was to see if there were effects of OI in this unique 2AFC paradigm for related and unrelated stimuli. The second goal was to observe if the amount of OI was differentially affected by the stimuli type in a case where bias was not a factor, with the prediction that there would be a greater amount of interference for related words. Using a 2AFC paradigm allows for a more accurate measure of discriminability. Thus, accuracy is a purer measure of discriminability, with no bias as there was in Experiment 1.

With the robust findings in the literature of OI, it was expected to see significant changes in accuracy across bins, which would be indicative of performance decreasing over the course of the test list with each successive bin. However, the ANOVAs on proportion correct between bins were only significant between bin one and bin four and only in the animal condition. Additionally, results from the linear regression model suggests that there was weak evidence of OI in the animal condition and no effects of OI in the emotion condition. In order to test the hypothesis that there would be significantly more OI for related words compared to unrelated words, the slopes of related and unrelated words for every individual were computed and compared. The insignificance of this analysis suggests that OI is equal for related and unrelated words, which is in opposition to the original hypothesis. The multilevel model analyses provide further evidence that there were weak effects of OI in the animal condition, no effects of OI in the emotion condition, and not a greater amount of OI for related words.

The only two differences between Experiment 1 and Experiment 2 were the presentation of words at test (single item recognition or 2AFC) and the levels of processing task during encoding. There is evidence in the literature that levels of encoding is independent of the effects of OI (Kilic, 2012). There are also studies that show standard forced choice paradigms still produce effects of OI (Criss et al., 2011). Therefore, it is possible that this unique presentation of test pairs, where items are presented one at a time and then together, may be effecting performance and discriminability. For example, forcing slower responses could lead to higher accuracy throughout the test. This question was the basis for Experiment 3, which is a replication of this experiment, but using a standard 2AFC paradigm.

Experiment 3

Procedure

Experiment 3 is a replication of Experiment 2, with the only difference being a standard 2AFC paradigm was used at test. Therefore, participants saw both words presented simultaneously at test, rather than one at a time and then together.

Participants

Recruiting for Experiment 3 was identical to the methods used in Experiment 2. Using the same cutoffs as Experiment 2, data from a total of 30 and 31 participants were included in the analyses for the emotion and animal conditions, respectively.

Hypotheses

Even though Experiment 2 did not show a greater amount of OI for related words, using this standard paradigm, it was again predicted that there would be a greater effect for related words with the theory being similar features would create more interference.

Results

Analyses for Experiment 3 were identical to those used in Experiment 2, so data were divided into the different stimuli type, either related or unrelated, for both the animal and emotion condition and by bins. To observe effects of OI, the mean proportion of correct responses was calculated for every participant. Figure 8 plots the overall proportion correct per bin for related and unrelated words for both conditions. A 2 (stimulus type) x 4 (bin number) repeated measures ANOVA was conducted for each condition separately. For the animal condition, there was only a main effect on bin number, $F(1,30) = 4.60$, $p = .005$. There were no main effects found for the emotion condition, not even for bin number, $F(1,27) = 2.17$, $p = .098$. The slopes of every individual were obtained for both related and unrelated words to see if OI differentially affects stimulus type. Performing a paired sample t test on the animal condition showed a significant difference between the related and unrelated slopes, $t(30) = -2.96$, $p = 0.006$. The same analysis was conducted for the emotion condition, but there was no significance, $t(29) = 1.03$, $p = 0.310$.

As in Experiment 2, a linear regression model was implemented to observe performance on a trial by trial basis. Using the individual data sets, the average proportion correct responses for every test trial (i.e. 1-96) was conducted and the lines of best fit for related and unrelated words were plotted. This again was done for both conditions (see Figure 9). From these results, the correlations between proportion correct and trial number were found. In the animal condition, the correlation for related words was -0.355 and -0.053 for unrelated words. This correlation was only significant for the related words, $t(94) = -3.69$, $p < .001$. For the emotion condition, the correlations were -0.152 and -0.303 for related and unrelated words, respectively. In this case, there was only significant effect on unrelated words, $t(94) = -3.09$, $p = 0.003$.

The same three multilevel models were used here as in Experiment 2. M_0, or the null model, had a fixed effect of stimulus type (i.e. related or unrelated) and random effect of stimulus type. To test the effects of OI, M_1 had fixed and random effects of test position to test effects of OI. Finally, M_2 observed the interaction between stimulus type and test position. Results are of the model comparisons are shown in Table 6. There were effects of OI in both conditions, but stronger in the animal. Additionally, there was also a marginal effect of an interaction between stimulus type and test position in the animal condition, suggesting that there was a greater amount of OI for the related words compared to the unrelated words.

Stimuli	Proportion Correct: Bin 1	Proportion Correct: Bin 2	Proportion Correct: Bin 3	Proportion Correct: Bin 4
Animal Pool				
Related	0.880 (0.16)	0.828 (0.22)	0.808 (0.13)	0.754 (0.13)
Unrelated	0.857 (0.15)	0.861 (0.21)	0.852 (0.13)	0.844 (0.13)
Emotion Pool				
Related	0.848 (0.14)	0.816 (0.22)	0.800 (0.19)	0.823 (0.14)
Unrelated	0.874 (0.16)	0.837 (0.21)	0.816 (0.18)	0.803 (0.19)

Table 5. Mean proportion correct responses for related and unrelated stimuli per bin for both emotion and animal stimuli.

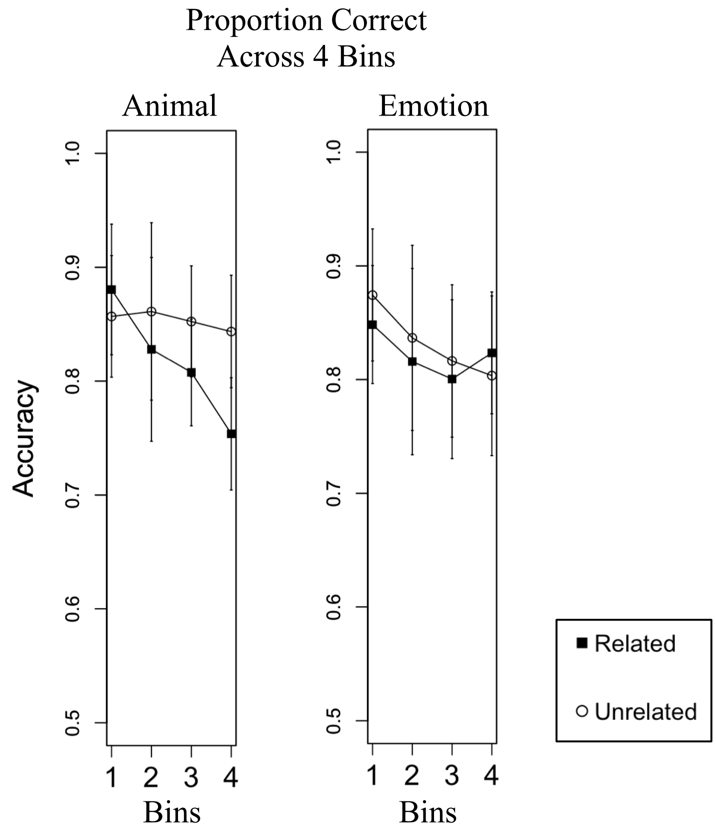

Figure 8. Average proportion correct for all related words and unrelated across four test bins of 24 trials each for animal condition and emotion condition.

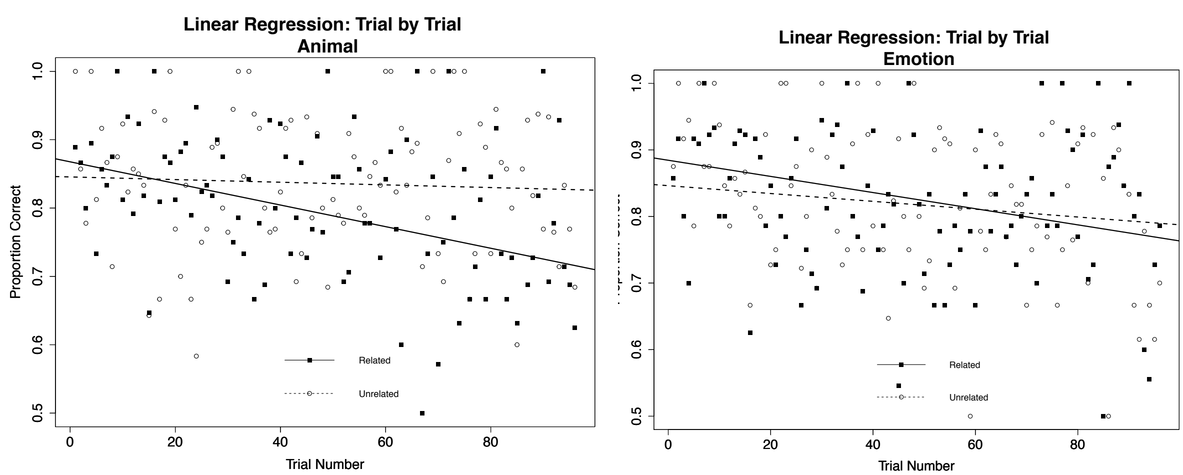

Figure 9. Trial by trial linear regression for proportion correct responses for all related words and unrelated. This includes the lines of best fit for both animal and emotion condition.

Model Animal	Fixed Effects	Random Effects	AIC	BIC	df	χ^2
M_0	Stimulus Type	Stimulus Type	2462.6	2492.2	2791	
M_1	Stimulus Type, Test Position	Stimulus Type, Test Position	2446.7	2500.1	2787	M_1-M_0 = 23.89**
M_2	Stimulus Type, Test Position, Interaction	Stimulus Type, Test Position, Interaction	2443.7	2503.0	2786	M_2-M_1 = 5.00*

Model Emotion	Fixed Effects	Random Effects	AIC	BIC	df	χ^2
M_0	Stimulus Type	Stimulus Type	2100.2	2129.2	2436	
M_1	Stimulus Type, Test Position	Stimulus Type, Test Position	2096.2	2148.4	2432	M_1-M_0 = 12.01*
M_2	Stimulus Type, Test Position, Interaction	Stimulus Type, Test Position, Interaction	2097.0	2155.0	2431	M_2-M_1 = 1.24

*Table 6. The model comparisons for accuracy for the animal (top) and emotion (bottom) conditions. AIC is the Akaike Information Criterion and BIC is the Bayesian Information Criterion. The significant differences between χ^2 values of each model pair at $\alpha = 0.05$ is represented by * and $\alpha = 0.001$ is represented by ** sign.*

Discussion

Experiment 3 was a replication of Experiment 2, with the only exception being how the words were presented at test. While Experiment 2 used a double presentation, Experiment 3 used the standard paradigm of simply having both words presented at once during the test phase. The primary interest here was to investigate whether the modified test presentation used in Experiment 2 masked the effects of OI.

Interestingly, the animal and emotion conditions had overall different patterns of behavior when looking at how performance changed over the course of the test list. These results showed no effects of OI in the emotion condition and only evidence of OI for related words in the animal condition. This is an interesting result because it is more evidence suggesting that OI may not be as robust as thought. Experiment 1 showed that emotion words elicited a greater bias to call words old. Items that share related features may result in a greater memory bias because the words are more confusable at test and harder to discriminate. This same idea may also predict that these items with related features will also elicit greater OI. However, results from Experiment 1 and Experiment 3 show emotion words have an old memory bias, but do not have greater effects of OI.

As predicted, there was a greater amount of OI for related words compared to unrelated words, but only in the animal condition. This is again surprising as this was not the case for the emotion stimuli. Additionally, the regression analyses showed further support for a significant amount of OI for related words, but not unrelated words, in the animal condition. However, the results were reversed for the emotion condition, such that only the unrelated words had a significant effect of OI. This is arguably the most interesting result from this experiment because although the animal condition was almost exactly as predicted, the emotion condition appears to have the reverse findings, such that unrelated words have a greater amount of OI compared to related words. This suggests it is important to analyze data for different categories separately.

General Discussion

Behavioral Analyses

The present study explored effects of item relatedness (i.e. related and unrelated words) and OI in recognition memory using animal names and emotion words. The focus was to determine if item relatedness (similarity) leads to increased OI. The primary method of analysis was dividing the data into four bins to observe if and how discriminability changes over the course of the test. Experiment 1 utilized a single-item recognition task and showed different effects of overall bias among related and unrelated words, but no differences in OI. Because bias may contaminate the measure of d' in the first experiment, the second and third experiments used a 2AFC for which bias is not a factor. Results from Experiment 2 showed weak overall effects of OI, but importantly there was no significant difference in OI for related and unrelated words. Experiment 3 mirrored Experiment 2 except that a standard test presentation was used. For animal names, there was a greater amount of OI for related words compared to unrelated words, consistent with the original hypothesis. However, this was driven by the fact that there was no OI for the unrelated words. For the emotion condition, there were no effects of OI.

While there were significant effects of OI in the single item presentation experiment, there were only marginal effects of OI in the animal and no effects in the emotion condition in the 2AFC experiments. Furthermore, OI was not greater in the related stimuli compared to the unrelated stimuli across the board. Given the different word types used (animal or emotion) and the different testing conditions (single-item vs 2AFC), it is clear that the lack of stronger OI is not specific to one category of words or testing protocol. Thus contrary to expectation, the presence of shared features in related words did not lead to stronger interference effects at test.

However, arbitrarily dividing the data into four bins may not be the best approach for analyzing the data at hand. For example, trial 24 would be part of the same bin as trial 23, but not 25, when in reality the difference between trials 24 and 23 should be the same as 24 and 25. Therefore, two other methods were used to analyze the forced choice data. First, a trial by trial linear regression was used to see what is happening across every single trial. Consistent with the previous analysis, these results showed that in general there was not a difference in the amount of OI for related compared to unrelated stimuli. However, just by looking at the graphs one can see the incredible amount of variability in the data across trial number. For that reason, the second analysis used was a linear mixed effect model was used. This is arguably the best approach because it accounts for every trial as well as random effects (e.g. subject variability). In both forced choice tasks, these results showed OI for the animal condition, but weak to no OI for the emotion condition. Again, in general OI was not greater for related compared to unrelated words, but did show some marginal effects.

This work started off with a purely conceptual theory, such that the overlapping features in related words would drive a greater amount of OI in those same words across a test list. Taking these three experiments together, this study failed to show evidence that OI is stronger for related than unrelated words. Nonetheless, these results raise interesting questions about the nature of item and testing effects, and more broadly speak to the importance of considering item characteristics, test position, bias, and discriminability when trying to investigate memory processing. These effects are discussed next in terms of implications for theories of memory processing.

Item-Noise Model Implications

Item-noise models of recognition memory make certain predictions about the specific mechanisms driving interference. Retrieving Effectively from Memory (REM) is one example of an item noise model that assumes interference is due to both similar context information and similar item information (Clark & Gronlund, 1996; Criss et al., 2011; Gillund & Shiffrin, 1984). From a theoretical standpoint, items from the same category share many common features, making them more confusable in memory at test. Item noise models would predict that stimuli with overlapping features would result in a greater bias in single item recognition. These models would successfully predict an increase in hit rates and false alarm rates for the negative emotion words as shown in Experiment 1. This same idea can be transferred to understanding OI, such that item relatedness would result in more confusability, creating a larger effect of OI. It is unclear whether these models can simultaneously account for an overall bias without stronger OI. Future work will be needed in this area.

Output Interference: Study and Test Context

It is interesting that there were significant effects of OI in the single item task (Exp. 1), but weak or no effects of OI in the 2AFC tasks (Exps. 2 and 3). One possible explanation for this is that the 2AFC tasks have sufficiently different study and test contexts to prevent OI from occurring. Although there is evidence of OI in forced choice tasks (Criss et al., 2011), the unique 2AFC paradigm used in Experiment 2 had a significantly longer test presentation. During the study phase participants see one item at a time, while at test they see one item, a second item, and then both together. The differences in study and text context in Experiment 2 could potentially explain the lack of overall OI. When applying to the standard 2AFC presentation, which arguably has a more similar study and test context, there are greater effects of OI. However, the effects were still not as strong as in single item recognition experiment when the study and test contexts were identical. Future studies should be conducted manipulating the study and test contexts to see the effects of OI in different settings.

Limitations

One limitation to this study is the limited number of stimuli. To successfully look at effects of OI for categorized words requires a long test list to observe the change in performance over the course of the test. However, it is difficult to create a large stimulus set with categorical words because there is a limited number of interrelated stimuli for a given category. Additionally, binning the data creates an even greater problem with the limited number of items. For instance, in the 2AFC tasks there were a total of 96 test item, but that was divided in half for related vs. unrelated words, then divided into four bins. This resulted in only 12 trials of data for each participant per bin. On top of that, some of those trials were excluded due to the cutoffs. This could explain why there was so much variability in the data when plotting accuracy per bid. Therefore, this suggests that binning data may not be the greatest approach because it involves dividing the data rather arbitrarily and leaving few trials of data. Using a linear regression model to observe performance trial by trial gives a more accurate measure and understanding of how performance changes across the test, but a multi-level model appears to be the best because it can also take random effects into account.

Additionally, future studies should explore the effects of OI in other categories besides animal names and negative emotion stimuli. Although the emotional words used in this study satisfied the goal of higher relatedness, there are other potential effects related to emotional valence, arousal, and dominance that could have influenced the present results (Adelman & Estes, 2013; Doerksen & Shimamura, 2001; Dougal & Rotello, 2007). It would be interesting to compare other categories of related items (e.g., household items) to determine how robust the effects of relatedness on OI are.

Conclusion

The present study shows mixed effects of OI on different types of stimuli in recognition memory. Although there were significant effects of OI in the single item presentation experiment, there were moderate or no effects of OI in the 2AFC studies. These findings suggest that effects of OI may not be as robust as thought. Surprisingly, in general there was not a greater amount of OI for related words compared to unrelated words. Thus these results suggest that item relatedness does not lead to stronger OI. It is unclear whether certain models of memory would be able to capture this counterintuitive behavior. Future work is needed to investigate whether memory models can account for this lack of OI and the interaction between item relatedness and the effects of OI.

Appendix A
Preregistration Form

As Predicted: *"Effects of Item Relatedness on Output Interference - Syracuse, 2016"* (#1819)

Created: 11/04/2016 08:32 AM (PT)

Author(s)
Jennifer Sloane (Syracuse University) - jfsloane@syr.edu
Corey White (Syracuse University) - cnwhite@syr.edu
Ryan Curl (Syracuse University) - rcurl@syr.edu

1) What's the main question being asked or hypothesis being tested in this study?
There would be different effects of output interference, decrease in performance over the test list, depending on the stimulus type (related vs. unrelated)

2) Describe the key dependent variable(s) specifying how they will be measured.
Correct responses will be analyzed and averaged for each participant

3) How many and which conditions will participants be assigned to?
Two conditions: a mixed list with half unrelated stimuli and half negative emotional stimuli and a mixed list with half unrelated stimuli and half animal names

4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
I will observe correct and incorrect responses across four bins, with 25 trials per bin. However, I will exclude the first trial from my analyses due to primacy effects. I will set a priori cutoffs to exclude participants whose performance are at chance (50%) or below and individual trials where the respond time is faster than 300ms or exceeds 3000ms. Then, I will calculate Bayesian t-tests and compare the slopes for related and unrelated stimuli. Additionally, I will conduct a sequential analysis, where I will separate trials by whether the preceding trial was related or unrelated stimuli to look at whether there are carryover effects of trial type

5) Any secondary analyses?
If there are no significant differences between the related items, I will aggregate the data and then run the same analyses for related and unrelated stimuli. If there are significant differences, I will run separate analyses for animal names and unrelated stimuli and for emotional and unrelated stimuli

6) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
I will collect data on 50 participants for both conditions

7) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)

8) Have any data been collected for this study already?
No, no data have been collected for this study yet

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. Cognition, 129(3), 530-535. doi:10.1016/j.cognition.2013.08.014

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain Activity Underlying Emotional Valence and Arousal: A Response-Related fMRI Study. Human Brain Mapping, 23(4), 200-209. doi:10.1002/hbm.20048

Aue, W. R., Criss, A. H., & Prince, M. A. (2015). Dynamic memory searches: Selective output interference for the memory of facts. Psychonomic Bulletin & Review, 22(6), 1798-1806. doi:10.3758/s13423-015-0840-5

Bates, D. M., M; Bolker, B; Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Buchanan, T. W., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In S. C. Moore, M. Oaksford, S. C. Moore, & M. Oaksford (Eds.), Emotional cognition: From brain to behaviour. (pp. 9-34). Amsterdam, Netherlands: John Benjamins Publishing Company.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin & Review, 3(1), 37-60. doi:10.3758/BF03210740

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11(6), 671-684. doi:10.1016/S0022-5371(72)80001-X

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

Cunningham, S. J., Brebner, J. L., Quinn, F., & Turk, D. J. (2014). The self-reference effect on memory in early childhood. Child Development, 85(2), 808-823. doi:10.1111/cdev.12144

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. Emotion, 1(1), 5-11. doi:10.1037/1528-3542.1.1.5

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. Psychon Bull Rev, 14(3), 423-429.

Francis, W. N., & Kucera, H. (1982). Frequency analysis of English usage. Houghton Mifflin; Boston.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. Memory & Cognition, 38(7), 833-848. doi:10.3758/MC.38.7.833

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

Glover, J. A. (1989). The 'testing' phenomenon: Not gone but nearly forgotten. Journal of Educational Psychology, 81(3), 392-399. doi:10.1037/0022-0663.81.3.392

Goh, W. D., & Goh, C. K. (2006). The roles of semantic similarity and proactive interference in the word length effect. Psychonomic Bulletin & Review, 13(6), 978-984. doi:10.3758/BF03213912

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

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. Trends in Cognitive Sciences, 5(9), 394-400. doi:10.1016/S1364-6613(00)01707-1

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. Journal of Verbal Learning & Verbal Behavior, 10(5), 562-567. doi:10.1016/S0022-5371(71)80029-4

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? Memory & Cognition, 31(8), 1169-1180. doi:10.3758/BF03195800

Kilic, A. (2012). Output interference and strength based mirror effect in recognition memory. (74), ProQuest Information & Learning, US. Retrieved from http://libezproxy.syr.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-99160-442&site=ehost-live Available from EBSCOhost psyh database.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25(2-3), 259-284. doi:10.1080/01638539809545028

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, US: Cambridge University Press.

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

Monaco, J. D., Abbott, L. F., & Kahana, M. J. (2007). Lexico-semantic structure and the word-frequency effect in recognition memory. Learning & Memory, 14(3), 204-213. doi:10.1101/lm.363207

Montefinese, M., Zannino, G. D., & Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. Psychological Research, 79(5), 785-794. doi:10.1007/s00426-014-0615-z

Nabeta, T., & Kusumi, T. (2009). False memories in the Deese-Roediger-McDermott (DRM) paradigm: Selective review of the production mechanism and phenomenology. Japanese Psychological Review, 52(4), 545-575.

Norman, D. A., & Waugh, N. C. (1968). Stimulus and response interference in recognition-memory experiments. Journal of Experimental Psychology, 78(4, Pt.1), 551-559. doi:10.1037/h0026637

Roediger, H. L., III, & 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., III, & McDermott, K. B. (2004). Creating False Memories: Remembering Words Not Presented in Lists. In D. A. Balota, E. J. Marsh, D. A. Balota, & E. J. Marsh (Eds.), Cognitive psychology: Key readings. (pp. 338-354). New York, NY, US: Psychology Press.

Schulman, A. I. (1974). The declining course of recognition memory. Memory & Cognition, 2(1-A), 14-18. doi:10.3758/BF03197485

Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. Nature Neuroscience, 7(12), 1376-1380. doi:10.1038/nn1353

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(2), 267-287. doi:10.1037/0278-7393.21.2.267

Starns, J. J., Chen, T., & Staub, A. (2017). Eye movements in forced-choice recognition: Absolute judgments can preclude relative judgments. Journal of Memory and Language, 93, 55-66. doi:10.1016/j.jml.2016.09.001

White, C. N., Kapucu, A., Bruno, D., Rotello, C. M., & Ratcliff, R. (2014). Memory bias for negative emotional words in recognition memory is driven by effects of category membership. Cognition and Emotion, 28(5), 867-880. doi:10.1080/02699931.2013.858028

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

Jennifer Sloane

334 Miles Avenue• Syracuse, NY 13210 • Phone: 410-299-9787
E-Mail: jfsloane@syr.edu • Teaching Portfolio: Jenniferfsloane.wordpress.com

Education

Graduate Student: Cognition, Brain, and Behavior Program, Syracuse University
Advisor: Corey White, Ph.D.

B.A. *Psychology*, University of Maryland College Park, 2014

Research Experience

Graduate Student, Brain and Behavior Lab — August 2015 - Present
PI: Corey White, Ph.D., Syracuse University

Research Assistant, Neurocognitive Development Lab — June 2012 – August 2015
PI: Tracy Riggins, Ph.D., University of Maryland

Research Assistant, "Hippocampal Memory Network" Project — January 2015 – August 2015
PI: Tracy Riggins, Ph.D., University of Maryland

Research Assistant, Decision, Attention, and Memory Lab — August 2014 – May 2015
PI: Michael Dougherty, Ph.D., University of Maryland

Teaching Experience

Teaching Assistant, Introduction to Psychology, Syracuse University — August 2015 - Present

Instructor, Cognitive Psychology, Syracuse University — July 2016 – August 2016

Tutor for Elementary Students through Women in Science and Engineering Program — October 2016 - Present

Teaching Assistant, Developmental Psychology, University of Maryland — August 2014 – December 2014

Tutor for student athletes, Academic Support and Career Development Unit, University of Maryland — January 2014 – August 2014

Volunteer and teaching assistant, University Park Elementary School, Hyattsville, MD — September 2013 – December 2013

Publications/Presentation/Reviews

White, C.N., Curl, R.A., Sloane, J.F. (2016). Using decision models to enhance investigations of individual differences in cognitive neuroscience. *Frontiers.*

Sloane, J.F., Curl. R.A., & White, C.N. Understanding effects of categorization on memory bias and discriminability as a function of output interference. Poster presented at the 2016 Mathematical Psychology Conference, New Brunswick, NJ

Curl. R.A. & Sloane, J.F. Characterizing the Relationship between Semantically Similar and Dissimilar Items in Recognition Memory. Poster presented at the 2015 Psychonomics Conference, Chicago, IL

Lumsden, J. A., Attwood, A. S., & Munafo, M. R. (2016). Comparison of Data Generated by Gamified and Standard Stop Signal Tasks. Reviewed Manuscript for PLOS ONE journal.

Honors and Awards

Women of Mathematical Psychology Travel and Networking Award of $800	August 2016
Member of Women in Science and Engineering	September 2016 - Present
Mark S. Harper Award for Excellence in Psychology, University of Maryland	May 2015
Member of Phi Beta Kappa	November 2015 - Present
Deans List, University of Maryland	
College Park Scholars Program, Public Leadership	
Primannum Honor Society	
The National Society of Collegiate Scholar	