Memory of Words: A Categorization Task

2018

Paulina Maxim

University of Central Florida

MEMORY OF WORDS: A CATEGORIZATION TASK

by

PAULINA MAXIM

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Psychology
in the College of Sciences
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Summer Term, 2018

Thesis Chair: Valerie Sims, PhD

ABSTRACT

Through the years, the Deese-Roediger-McDermott Paradigm has demonstrated to be a useful method of observing false memories from semantically related word lists. The present study was conducted fully online and measured memory performance dependent on categorization of words by using groups, as well as dragging words across the page as a form of interaction. In a 2 (Categorized, Non-Categorized) x 2 (Interactive, Non-Interactive) between-subject factorial experiment, 56 undergraduate students were shown 18 different lists of 15 associative words to be studied, one list at a time. Participants were given a free recall test immediately after studying each individual list. Participants also performed a recognition test after having studied and recalled all 18 lists, which consisted of 216 items; half of the words were presented throughout the studied lists, and the other half consisted of the 18 critical lure words as well as several other distractor items from a subset of word lists. It was hypothesized that participants in both the categorization and interaction condition would show the highest levels of accurate memory recall and recognition compared to those who were simply given a list to review. Findings did not support this hypothesis indicating no clear differences between participants who categorized (or not) or interacted with the lists (or not). High probabilities were found for words ranked as highly falsely recalled and low probabilities were found for those ranked towards the bottom, much like the findings in Stadler et al., (1999).

ACKNOWLEDGMENTS

This work could not have been done without the endless guidance and dedication of Dr. Valerie Sims and Dr. Mustapha Mouloua, who's pivotal role in advising me throughout the course of my thesis and throughout my undergraduate career, has successfully prepared me for my journey as a graduate student.

I would also like to thank my parents, Walter and Beatriz Maxim, without whom I would not be where I am today in life. Your love and support know no bounds and I cannot thank you enough for your encouragement towards my educational and career endeavors.

TABLE OF CONTENTS

INTRODUCTION 1

METHOD 10

Participants 10

Materials 10

Design 12

Procedure 12

RESULTS 14

DISCUSSION 17

APPENDIX A: RECALL WORD LISTS 21

APPENDIX B: RECOGNITION WORD LISTS 25

APPENDIX C: ORDER-OF-OPERATIONS PROBLEMS 30

APPENDIX D: DEMOGRAPHICS QUESTIONNAIRE 33

APPENDIX E: INFORMED CONSENT 36

APPENDIX F: IRB APPROVAL LETTER 39

REFERENCES 41

INTRODUCTION

Though James Deese's early work studying verbal intrusions did not spark wide interest in its time, it was not long before his work on false recall and recognition of words was brought to light. When Henry L. Roediger III and Kathleen B. McDermott conducted their experiments in 1995, the Deese-Roediger-McDermott Paradigm became an established model of understanding how extra-list intrusions become prevalent in recall related tasks. Several researchers have taken Deese's original work and modified the paradigm by manipulating word lists, critical lures, recognition tasks, and participant instructions (Roediger & McDermott, 1995; McDermott & Roediger, 1998; Gallo, Roberts & Seamon, 1997; Stadler et al., 1999). With each modification, more and more findings have added to the conversation in hopes of better understanding this contextually-based illusion.

In 1959, James Deese sought out to explain how the occurrence of extra-list intrusions could be accounted for based on the associative context of the word lists themselves. In his study, Deese used word association norms to develop lists that could evoke extra-list intrusions as well as measure the frequency of their occurrence. Thirty-six prior recall lists made up of 12 associative word items were used as the stimuli in conducting these association tests. Subjects were instructed to write down a word that came to mind next to each corresponding stimulus item. The results of this word association task were used in comparison to the frequencies from the Minnesota norms based off the original Kent-Rosanoff word association test (Russell & Jenkins, 1954). A strong correlation indicated that "the variance in probability of intrusion [was] determined by variation in mean association strength" (Deese, 1959).

Free association has demonstrated its strength in predicting word intrusions in immediate recall. Roediger and McDermott set out to observe these intrusions in not just the single-trial free recall paradigm, but in recognition tasks as well (Roediger & McDermott, 1995). In their first experiment, subjects were tested for false recall of studied words and were later tested for false recognition of both the studied and non-studied items. Unlike in Deese's experiment, only six lists with the highest intrusion rates were chosen: *chair, mountain, needle, rough, sleep*, and *sweet*. Studied items were selected from the Russel-Jenkins word association norms (1954). The non-studied words used in the recognition test were chosen on the bases of three factors – the critical lures corresponding to each studied list, a few weakly related items from the original association norms, and several unrelated items to the ones from the six lists chosen for the experiment. For the recognition test, subjects were presented a mixture of the studied and non-studied items per each list in a block format (see Underwood (1965) and McDermott (1996), for discussion on block/random presentation). The non-presented critical lure was recalled 40% of the time, similar to the rate of words found in the middle of a studied list. Given the like rate of recall between studied and non-studied items, Roediger and McDermott were able to clearly demonstrate a false recall trend with the paradigm.

In addition to determining false memory, Roediger and McDermott obtained subject confidence ratings to measure how confident subjects were in having seen a particular word in the previous six lists presented. Though the present study does not test for confidence ratings, it is important to point out that during the recognition test in 1995, accuracy of item recognition was quite high, with an 86% hit rate and as low as a 2% false alarm rate (Roediger & McDermott, 1995). However, when observing individual confidence accuracy ratings for the studied versus the non-studied items, subjects rated being “sure” of having studied critical lures that were never even presented. Roediger and McDermott incorporated the use of “remember” and “know” judgments in the recognition test; a measure of high and low confidence ratings meant to capture a subject’s phenomenological experience (Tulving, 1985).

For the purpose of the present study, we will not be going into great depth involving these judgments, though we will explore the potential differences between varying conditions. These include learning a word list and immediately recalling it prior to performing a recognition test, versus learning a list and skipping straight to the recognition test while never having recalled the items at all. Roediger and McDermott explore this idea by having their subjects study 24 word lists of 15 associates each. The researchers conducted a within-subjects experiment in which subjects studied two thirds of the lists, half of which items were recalled immediately and the other were not. The remainder eight lists were not studied at all. Subjects were instructed to use old/new and remember/know judgments and were tasked with recognizing 48 studied words from various serial positions and 48 non-studied words consisting of unseen critical lures plus several items from the lists not presented to the subjects. Researchers found that the non-presented critical lures were recalled for over half of the lists (Roediger & McDermott, 1995). Results for the recognition portion of this experiment are fascinating in the proximity of hit rates and false alarm rates. Per their findings, there was such a minimal gap in correct judgments for critical words and non-presented words, it was determined that subjects had a difficult time distinguishing between list items overall.

Understandably, the “study + recall” condition yielded the highest overall rates, particularly for ‘remember’ judgments, indicating a significant effect of recall memory on the subsequent recognition memory. Findings show that false recognition of studied items may be instigated at the time of encoding (Underwood, 1965). The theory poses that subjects may be coming up with an associate word pair for the studied item, such as seeing the word *hot*, but thinking about the word *cold*. Later during the recognition task, if presented with the word *cold*, subjects are more likely to claim having seen the word when it was never encoded as a studied item (Roediger & McDermott, 1995). Other theories suggest that the activation of the associate critical item is implicit and false recall and recognition may occur even without the subject being aware of this activation. This internal source is understood through "reality monitoring," a process of remembering information through reasoning, imagination, and thought (Johnson & Raye, 1981). This activation signifies the availability of the critical lures that were never studied. The term "superadditive priming" is used to explain this availability, since lure words were responded to more quickly than studied list items (Hancock et al., 2003). Information to make source attributions comes from these sources and other criteria such as "semantic plausibility" (Mather, Henkel, & Johnson, 1997). Other claims argue that a leading cause for poor performance on a recognition test could be in part due to the delay in time between studying multiple lists and completing the recognition test itself (McDermott, 1996). Researchers found proportional differences between initial tests and free recall two days later, but did not find any significant differences between immediate recall and a 30 second delay. Furthermore, the position of a critical item in a subject's recall output was not as affected as was the production levels of the list items themselves (1996). Stronger findings regarding output position for the critical item show that over half the time, targets are found in the last fifth of the subject's recall output (Roediger & McDermott, 1995).

The general idea behind the DRM paradigm, is for the intrusion word to be unknown or unseen prior to recall. Although this was not the case in Deese's original study, he argues that a single presentation of the word followed by a single recall trial would not lead to a high frequency of intrusion, as is demonstrated by weak correlations (1959). In experiments conducted by McDermott and Roediger (1998), subjects were shown the critical target in half of the studied lists in various serial positions throughout. When tested for recognition accuracy, subjects were asked to rate whether they were sure the word was old ('Presented') or new ('Nonpresented'). The findings for this experiment (Experiment 1, Figure 1), should depict opposite-like bar graphs for each condition, though not the case. Response probabilities for presented critical items are clean and accurate at a 79% hit rate, while probabilities for non-presented targets totaled 64% of false-alarm judgments of recognizing the critical word as old (McDermott & Roediger, 1998). In Experiment 2, researchers placed a one-item recognition test immediately after the studied word list. This time, slightly more accurate response ratings were found with a hit rate on 80% of the trials, though false alarm ratings still quite high at 38%.

McDermott and Roediger (1998) not only found a way to test direct memory of a presented and non-presented lure, but found a possible alternative in avoiding the confound of delay on memory recall as was discussed in McDermott's earlier experiment (1996). These findings encompass the former two theories regarding a subject's performance on the recognition test. Researchers initially put into question the effects of having first recalled a list on how accurate the list would be recognized, and later the effects of time delay. By immediately responding to the recognition test, subjects did not experience the initial recall that could have potentially interfered with recognizing studied words nor did time play a significant role in the retrieval of the studied information.

In staying true to the established methods of the DRM paradigm, most research incorporates both recall and recognition tests in hopes of discovering new information. Variations of the original method have led to new findings, but an element that has demonstrated itself to be a key component is the composition of the word list itself. Each word list has been carefully selected throughout the years based on qualities such as the strength of the associate terms (i.e. forward and backward associations), the list length, and whether the critical item is included (Deese, 1959; Roediger & McDermott, 1995; McDermott & Roediger, 1998). Compared to Deese's study in 1959, Roediger and McDermott used associated word items in each list, lending higher reports of high false recall and recognition. Deese's results are not as comparable since some of the recall lists did not contain associate items that produced the critical target words. For example, critical target words such as *cold* and *sleep*, are composed of associate items that yielded the critical targets, unlike the list for *butterfly*, whose word list did not exactly generate the target word, thus not eliciting reports of false recall (Deese, 1959). Gallo and Roediger (2002) demonstrate these findings in their experiment using lists constructed of weak associates, rendering low levels of false recall. Much like Roediger and McDermott's experiments, the present study uses longer lists of related words with the intention of inducing higher false recall and recognition frequencies.

The current study was primarily modeled after Stadler, Roediger, and McDermott's (1999) experiments attaining normative data for 36 lists consisting of 24 from Roediger and McDermott (1995) and 12 additional others unofficially published by McDermott (1995). Subjects saw 15 associate items in order of strongest to weakest associate strength. The recognition test presented 108 words, 54 studied items, 18 critical targets, and 36 words unrelated to the studied lists. Stadler and colleagues sorted the recall data into the top 18 lists that elicited the highest amount of false recall and recognition and the bottom 18 lists that produced the lowest. For example, the critical lure, *window*, was falsely recalled by 65% of the subjects, while only 10% of the subjects recalled the lure word, *king* (1999). Similar studies have used this method of utilizing top associates alongside their corresponding associative strengths (Hancock et al., 2003).

For the present study, forward and backward associative strength values were used to determine which words would be chosen to incorporate into the recognition test later modeled in the method section. Roediger and colleagues (2001) define forward associative strength (FAS) as the strength of the associate connection of the critical item to its respective associate list and backward associative strength (BAS) as that same connection, but from the associate words to the critical target. Researchers have shown that high associative strengths typically cause false recollection rejection rates to drop because it becomes more difficult to resist accepting the critical lures. It has also been shown that increasing the BAS shows no effect on accepting critical distractors on the basis of semantic overlap with target words alone (Brainerd & Wright, 2005).

The present study calls into question whether there are specific strategies used by subjects in studying word lists. As later illustrated, this paper observes categorization as a method of actively learning information and minimizing the frequency of false memories. There is limited research on the benefits of categorization as a memory tool. Previous research has observed the implication of semantic categorization on short-term memory (Wickens, 1970). Wickens discusses the occurrence of "proactive inhibition," which is considered to be the interference that occurs in learning subsequent information. By using short-term memory techniques such as "taxonomic categories," humans can better retain information based on the effective encoding that initially took place. Such categories are typically based on semantic classification, which has been shown to be more effective than classifying the physical characteristics of a word, for instance (1970). Other researchers have studied the use of conceptual categories in facilitating free recall (Kroes & Libby, 1971). Researchers found that in stronger conceptual categories, clustering tends to occur at a higher rate. Furthermore, clustering "non-exhaustive" categories using more versatile concepts such as clothing items, for example, is more powerful than clustering exhaustive categories based on limiting instances, such as the rhyming of two or more words (1971). Additional research has been conducted by observing the clustering of lists of high and low frequency associate items. Findings show that there are benefits of clustering using block presentation for word recall particularly amongst the high-frequency list items rather than the low-frequency ones (Cofer, Bruce, & Reicher, 1966). Researchers also determined that the duration of the presentation of an associate can lend to clustering and word recall as well.

Using categorization and conceptualization can play a meaningful role in learning and applying new knowledge. Cairncross and Mannion (2001), state that deep learning occurs when active engagement occurs, such as "learning by doing" or putting new knowledge to practice. In other words, interacting with the information in some way or another. This engagement was observed through the medium of online learning. For the present study, the researcher plans on observing human-computer interaction in the form of dragging words as the main source of interaction. This interaction is operationally defined as the physical dragging of a list item from one point to another (i.e. the left side of the screen to the right side), while involving computer accessories, like a mouse. Via categorization and interaction, it is hypothesized that participants who are encouraged to categorize associate list items will have more accurate recall and recognition than those who do not categorize. Likewise, participants who physically interact with the word lists on the computer, will show better memory performance and less overall false recall and recognition. Traditionally, the DRM paradigm has been tested in a controlled laboratory setting, recording each individual word and playing it out loud for participants to hear (Dees, 1959; Roediger & McDermott, 1995). Due to an already developed methodology, most research has not deviated from this norm. What makes the present study unique is that not only does the researcher intend on determining the effects of categorization on false recall and recognition, an application rarely seen in the memory research, but proposes a variant in the method by conducting the research fully online, expanding on the literature for modern-day learning mediums.

METHOD

Participants

The participants in this study consisted of 56 undergraduate students from the University of Central Florida. All participants were recruited through *Sona*, a UCF Psychology Research Participation System. Participation was voluntary, though participants were given Sona credit to be used toward their academic courses. Approximately 71 participants attempted to participate in this study, though only the 56 gave their consent and completed the research in its entirety.

Materials

Eighteen lists were tested (See Appendix A). All 18 lists were obtained from a set of 36 lists based on norms developed by Stadler, Roediger, and McDermott (1999). Of the 18 lists, 16 derived from Roediger and McDermott's original study (1995) and the remainder from McDermott (1995). The 18 lists tested in the present study consisted of 15 associates items, or words of a critical target word meant to be falsely evoked (Stadler et al., 1999). The lists tested were chosen based on the proportion of participants who recalled a critical item (target word) based on the 15 associates shown from each list. Eight of the top most recalled critical items and eight of the bottom least recalled critical items were chosen to test the effects of false recall of unseen critical target words (1999). Several standard order-of-operations math problems were also presented to the participants before each subsequent word list (See Appendix C).

An additional subset of words and their respective critical items were used to test recognition of words seen and not seen throughout the recall phase of the study (See Appendix B). This subset was obtained from a set of 55 compiled word lists found in a study conducted by Roediger and colleagues (2001). Eighteen critical items and a few of their corresponding associates were chosen to serve as unrelated distractor words in the recognition task in testing accurate recall of the words actually presented to the participants. The 18 lists tested in the initial recall phase (Stadler et al., 1999) can also be found amongst the 55 lists compiled by Roediger et al. (2001).

A total of 216 words were presented to the participants for the recognition task; 108 words came from the 18 lists presented during the test phase and the remaining 108 words came from the additional subset that were never presented to the participants. The words seen by participants were chosen based on their backward associative strength (BAS). The BAS determines the strength of the association of associative words to the corresponding critical items (Roediger et al., 2001). Thus, the top six associates with the highest BAS values were selected for all 18 studied lists at an attempt of eliciting false recall of the critical items. Furthermore, the words not seen by participants, aside from the 18 critical targets, were chosen based on how closely they could potentially evoke an illusion of having seen a word during the study phase. These words were chosen on a superficial basis and were not tested in developing additional norms for the subset list.

A basic demographics questionnaire (See Appendix D) was created for participants to fill out at the end of the study containing basic research questions such as age, biological sex, gender, ethnicity, and race. The participants were also given an informed consent form (See Appendix E) prior to the start of the study detailing information regarding the purpose of the study, the procedure, and the contact information of the research team. Given the online nature of this study, there were no specific tools and technology used. Participants were at liberty to use any electronic device (i.e. desktop or laptop computer, tablet, smart phone) with access to the internet accompanied by the Qualtrics study link via Sona.

Design

A 2x2 between-subject factorial design was used in the present study. All 56 participants were randomly assigned to one of four conditions. Condition 1, Non-Categorized/Interactive, forces the participant to interact with the word list by dragging each word to a box on the right side of the list, though not having the option to categorize the words. Condition 2, Categorized/Interactive, forces the participant to interact with the words by dragging each word into as many or as few categorical groups using the six empty boxes provided. This allows the participant to categorize the words to his or her liking in hopes of using categorization as a tool to better remember the word list. Condition 3, Non-Categorized/Non-Interactive, does not involve categorizing the list nor interacting with it. The participant is simply asked to read and review the list presented. This condition can be considered a control to better measure the effects of categorization and interaction with the lists of words. Condition 4, Categorized/Non-Interactive, allows a participant to group each word into categories by labeling, as opposed to dragging, using as many or as few of the numerical values 1 – 6.

Procedure

Once the participants were assigned to a condition, they were given thorough instructions of how to complete the task, as well as a brief practice of the task for either dragging or labeling. No explicit instructions were given regarding the nature of the study involving false memory, nor were participants warned of the probability of committing false recall and recognition. Regardless of the condition, all participants were shown the same 18 word lists in random order. In the study phase, participants were given 60 seconds to review the 15 list items presented on the screen. Once the time ran out, participants immediately entered the recall test, in which they were instructed to type out every word that they could recall from the previous list in 60 seconds. This study and recall procedure was conducted for all 18 lists. Between each list, participants were given a mathematical distractor task, consisting of five simple standard order-of-operations problems in 30 seconds. These problems were presented with the purpose of minimizing the chance of interference of the previous list in subsequent learning. The duration of the experiment was approximately one hour, similar to the length of previous studies conducted using the DRM paradigm.

After having completed the recall portion of the study, participants were given a recognition test of 216 items consisting of 108 words that they had just been shown throughout the 18 lists as well as 108 words never before seen (the 18 critical targets for the lists shown are included in this number). Participants were instructed to state whether they recalled having seen the word on the screen in any of the previous lists presented to them by marking “Yes” or “No”, pressing the ‘f’ and ‘j’ keys respectively. All 216 words were presented in random order. Only the 18 critical target words of the 216 recognition items were analyzed for false recognition in order to directly compare them to their corresponding false recall rates from earlier in the study.

RESULTS

The data collected from the participants were coded and entered into SPSS for the four conditions: Non-Categorized/Interactive, Categorized/Interactive, Non-Categorized/Non-Interactive, and Categorized/Non-Interactive. A 2x2 between-subjects analysis of variance was run and the means and standard deviations were found for each condition. The calculated F values were obtained for both categorization and interaction groups as well as the interaction of the two independent variables together. This study did not show a significant effect of categorization on accurate memory recall [$F(1,52) = 1.37$, *n.s.*, $\eta^2 = .026$], false recollection of critical items [$F(1,52) = 2.48$, *n.s.*, $\eta^2 = .046$], nor false recognition of critical items [$F(1,52) = 1.02$, *n.s.*, $\eta^2 = .019$]. Findings also showed that there were no main effects found for the interaction variable on accurate memory recall [$F(1,52) = 2.04$, *n.s.*, $\eta^2 = .038$], as well as no effect on false recall [$F(1,52) = .44$, *n.s.*, $\eta^2 = .008$] nor false recognition [$F(1,52) = .76$, *n.s.*, $\eta^2 = .014$]. Lastly, there were no significant interaction effects between the independent variables on any of the dependent variables (accurate recall [$F(1,52) = .05$, *n.s.*, $\eta^2 = .001$], false recall [$F(1,52) = .03$, *n.s.*, $\eta^2 = .001$], and false recognition [$F(1,52) = .54$, *n.s.*, $\eta^2 = .010$]. It is believed that these findings are in part due to the small sample of participants recruited for the experiment.

A regression analysis of the variables showed the highest R Squared for accurate recall (.07) with the interaction variable showing the largest observed power (.288). For false recall of the critical item, R Squared was .05 and the largest observed power was found for the categorization variable (.339). For false recognition of the critical items, the R Squared was lowest at .04 and the power values were about the same with categorization coming in highest at .168. Based on this analysis, not much can be said about the true explanation for the findings. It can only be argued that approximately 7% of the findings are attributable to the effects found. Such a small value is disappointing, but despite the findings, there is a possibly note-worthy trend showing larger significance for the categorization variable for both false memory tests. With more participants, allowing for a larger effect size, it would be possible to determine if this value could continue to increment in significance.

As predicted, the categorized/interactive condition yielded the highest number of words recalled from the studied lists (M=173.44, SD=37.52) and the non-categorized/non-interactive condition resulted in the lowest amount of words recalled (M=147.00, SD=46.45). As an overall measure of recall accuracy, the categorization conditions outdid the interaction conditions by a mean difference of about 16. Though not statistically significant, a very clear observation can be made between the mean values of the false recall and false recognition tests. False recall across conditions was low, averaging at about 3.20. Strikingly, the mean for false recognition of the 18 critical items was 13.29. This increase in false memory can be seen for every individual who took both tests. It did not seem to matter the strategy behind the encoding of the lists, because these drastic jumps were found across all 56 participants. Age was a variable considered in understanding these changes in memory. One participant was initially considered an outlier (age 57) in a pool of fifty-five other college-aged students. After reviewing that individual's results, it was determined that there were no significant differences in memory recall and recognition attributable to age, and therefore the outlier was not removed.

One of the ways in which the critical target items were chosen, was based on Stadler and colleagues' (1999) word norms. The results found for the proportion of falsely recalled words in the present study closely resemble that of the top and bottom lists presented by the researchers.

For example, the word *king* is the critical item least likely to be falsely recalled, and indeed, only one participant recalled this item during testing. Additionally, top falsely recalled words per the 1999 norms, were also top items for the current study. Items such as *window*, *chair*, and *soft*, had the highest probability of false recall, with *soft* having a 41% rate of false recall.

DISCUSSION

Initial predictions indicated that categorization as a learning strategy alongside interaction with words, would lead to more studied words recalled and less overall false memory of critical lures. Though no statistically significant data could be retrieved from the results, case-by-case findings can be observed across all 56 participants. Generally, categorization conditions yielded higher recall, but most of which were false memories. An interpretation of these results is that categorization allowed participants to group the associate items well enough to recall them afterward, but perhaps in doing so, might have subconsciously added to these groups words that could "fit-the-mold" so to speak. It also took more time to categorize the word list and therefore time is an extraneous variable here. These finding pose the question of whether categorization could actually hurt the memory process rather than benefit it. The variable of interaction showed most of its effect on the amount of accurate words recalled out of the 270 that were presented. This additional finding could potentially demonstrate the beneficial effect of interacting with the word list by dragging, at least in an online setting. That is why this variable is so important. As online programs and e-learning begin to play a more prevalent role in education, it is important to measure the effectiveness of human-computer interaction. By utilizing tools such as dragging, tapping, or even as simple as highlighting, we can learn about the effectiveness of such strategies and apply them towards improving memory in a computerized environment.

Individual differences were carefully observed and were mostly found within the recall test of the experiment. For the categorization conditions, the researcher found that not every participant groups all words using the same characteristics. Some participants categorized the word lists using semantics, while others alphabetized the list items. Moreover, linguistic differences may have affected the way in which a participant categorized the words. All participants were told that the words were common nouns in the English language. How would one perform if not an English speaker? Though the present study did not ask for a participant's self -rating of English proficiency, it does question whether language affected recall performance and suggests theoretical implications in memory and language.

Cultural differences can also be considered influential factors in individual memory performance. Several observations found in the results may detail these differences. For example, in four separate cases, participants typed out a variation of the word "wooden" as a word they believed to have studied in the list for the critical item, *window*. It is possible that these participants thought of the word "wooden" more readily than others, based on their perception of what a house looks like. Another instance showed participants typing out the word "pizza" for the critical item *bread*, though never presented. In this case, perhaps culture plays a role in the type of food participants are more accustomed to eating. For instance, a participant who eats pizza as part of his or her diet (a food item typically consumed by a college student), might be quicker to pair the presented associate item "crust" to pizza rather than baguette, for example.

A way of questioning the significance of the results, is whether interference across word lists has any effect on participant performance. Researchers such as Benton J. Underwood, elaborate on the idea that interference effects on memory could be caused by a subject having previously learned a list before. If this is the case, it appears distributed learning is more beneficial than mass learning. The opposite is true if the subject has never studied the lists before, thus there is an equal or possibly better chance for retention since there should be no prior interference (Underwood, 1955; 1957). For the current study, participants were not shown the lists prior to studying them, therefore “mass learning” was the better option for review. However, approximately 45 minutes were spent learning 18 separate lists, and perhaps should have been more distributed in this case. Other findings suggest that having initially recalled items can affect recognition performance (Roediger & McDermott, 1995). As earlier discussed, it was believed that false recall later enhanced false recognition, which in their case resulted in “remember judgements” for the critical lures, but results for the present study do not specifically point to this theory. Instead, false recognition increased three-fold, which means that decent performance on the recall test should have led to slightly less accurate performance in the recognition test.

Given the results of this study, there is extensive work that can be done in finding statistical significance. There were several variables believed to have caused a high tendency of attrition in the present study. First, the experiment was not conducted in a controlled and supervised lab setting, therefore participants were more likely to not take the research seriously. In addition, there is a possibility that questions may have come up regarding experimental tasks, lending to participants having low confidence throughout the study. Lastly, the length of the study, though similar to that of previous work, is lengthy, resulting in a high incompletion rate. Furthermore, having participants complete the study on the same computer style could allow for more accurate interaction variable results in regards to the physical task that is done when interacting with the computer tools (i.e. mouse). Though the Qualtrics platform presents the study in the appropriate view according to the device type, the task of dragging is simply not the same if done on a tablet, for instance. Another limitation was present during the analysis of the experiment. The researcher was unable to view the order in which each word list was presented to each participant. Thus, it was impossible to calculate the rate of fatigue and its effect on recalling the first list compared to a participant's performance on recalling the eighteenth list. Previous DRM research does not typically discuss fatigue effects, but perhaps it should be considered in future research in conjunction with memory load as a contributing factor. These limitations must be addressed in order to see larger effect sizes and ultimately, a significant set of data.

Given the theoretical and practical implications of this study toward the field of education and cognition, it is important that this research be compared to the alternative learning platform of technology-free learning. Discerning the benefits of online versus in-person learning, is just one way of applying this research into a real-world setting. Categorization, interaction and other strategic methods of learning may bring up the question of dual-task interference in getting in the way of the primary task of remembering the words presented. A theoretical claim can be made on whether we should be encoding to *learn* and not solely remember. This study, along with other DRM experiments conducted in the past, can be used to measure how well participants can retain new information and for how long, implicating human working memory capacity across short and long-term storage. As the modern-day world becomes more technologically dependent, research that is set out to improve human-computer interaction will become more beneficial. The role that memory plays in learning is pivotal, and though the purpose of an education will never change, the way in which we educate will.

APPENDIX A: RECALL WORD LISTS

The 18-word lists and their 15 associate critical items

Anger:
mad
fear
hate
rage
temper
fury
ire
wrath
happy
fight
hatred
mean
calm
emotion
enrage

Bread:
butter
food
eat
sandwich
rye
jam
milk
flour
jelly
dough
crust
slice
wine
loaf
toast

Chair:
table
sit
legs
seat
couch
desk
recliner
sofa
wood
cushion
swivel
stool
sitting
rocking
bench

Army:
Navy
soldier
United States
rifle
Air Force
draft
military
Marines
march
infantry
captain
war
uniform
pilot
combat

Car:
truck
bus
train
automobile
vehicle
drive
jeep
Ford
race
keys
garage
highway
sedan
van
taxi

Cold:
hot
snow
warm
winter
ice
wet
frigid
chilly
heat
weather
freeze
air
shiver
Arctic
frost

Doctor:
nurse
sick
lawyer
medicine
health
hospital
dentist
physician
ill
patient
office
stethoscope
surgeon
clinic
cure

Girl:
boy
dolls
female
young
dress
pretty
hair
niece
dance
beautiful
cute
date
aunt
daughter
sister

Man:
woman
husband
uncle
lady
mouse
male
father
strong
friend
beard
person
handsome
muscle
suit
old

Fruit:
apple
vegetable
orange
kiwi
citrus
ripe
pear
banana
berry
cherry
basket
juice
salad
bowl
cocktail

King:
queen
England
crown
prince
George
dictator
palace
throne
chess
rule
subjects
monarch
royal
leader
reign

Mountain:
hill
valley
climb
summit
top
molehill
peak
plain
glacier
goat
bike
climber
range
steep
ski

Music:
note
sound
piano
sing
radio
band
melody
horn
concert
instrument
symphony
jazz
orchestra
art
rhythm

Soft:
hard
light
pillow
plush
loud
cotton
fur
touch
fluffy
feather
furry
downy
kitten
skin
tender

Thief:
steal
robber
crook
burglar
money
cop
bad
rob
jail
gun
villain
crime
bank
bandit
criminal

Needle:
thread
pin
eye
sewing
sharp
point
prick
thimble
haystack
thorn
hurt
injection
syringe
cloth
knitting

Sweet:
sour
candy
sugar
bitter
good
taste
tooth
nice
honey
soda
chocolate
heart
cake
tart
pie

Window:
door
glass
pane
shade
ledge
sill
house
open
curtain
frame
view
breeze
sash
screen
shutter

APPENDIX B: RECOGNITION WORD LISTS

Six words chosen per 18 studied lists based on backward associative strength (BAS).

mad
rage
temper
fury
ire
enrage

Navy
soldier
Air Force
military
Marines
infantry

butter
rye
dough
crust
loaf
toast

automobile
vehicle
drive
garage
sedan
van

table
seat
recliner
swivel
stool
rocking

hot
frigid
chilly
freeze
shiver
Arctic

nurse
physician
patient
stethoscope
surgeon
clinic

vegetable
kiwi
citrus
pear
banana
berry

boy
dolls
female
dress
pretty
date

queen
crown
throne
monarch
royal
reign

woman
lady
male
person
handsome
suit

hill
valley
climb
molehill
peak
climber

band
concert
symphony
jazz
orchestra
rhythm

sour
candy
sugar
bitter
honey
tart

thread
pin
thimble
haystack
injection
syringe

steal
robber
crook
burglar
rob
bandit

hard
pillow
loud
fluffy
downy
tender

door
pane
ledge
sill
curtain
shutter

Five words chosen per 18 subset lists to be used as distractor items.

command
obey
shout
halt
harsh

slow
stop
delay
traffic
speed

citizen
American
country
vote
patriot

carpet
rug
floor
room
wool

stove
cook
warm
oven
kitchen

black
white
night
blue
death

foot
ankle
arm
knee
mouth

cabbage
green
garden
patch
carrots

long
short
narrow
thin
underwear

justice
peace
law
government
supreme

shirt
tie
collar
jersey
cuffs

high
tall
jump
cliff
elevate

whistle
noise
tune
song
blow

pen
quill
felt
tip
cap

rough
smooth
bumpy
coarse
sand

cup
saucer
tea
coffee
drink

trouble
worry
danger
problem
police

cottage
home
cabin
fence
vines

The 18 critical items from the studied lists.

Anger
Army
Bread
Car
Chair
Cold
Doctor
Fruit
Girl
King
Man
Mountain
Music
Needle
Soft
Sweet
Thief
Window

APPENDIX C: ORDER-OF-OPERATIONS PROBLEMS

10 + 5 × (9 – 4)	4 + 9 ÷ 3 – 6	10 ÷ 2 × 4 – 7
(9 – 6 + 5) ÷ 2	2 × (8 – 4 + 7)	10 ÷ (6 + 4) × 9
(8 + 3 – 6) ÷ 5	8 + 6 × 7 – 3	3 × 4 – 8 + 5
10 × 7 + 3 – 4	(9 + 4 – 8) × 2	2 × 9 + 6 – 5
10 × (6 + 3) ÷ 2	10 + 2 × 3 – 8	9 – 4 ÷ 2 + 7
8 × 2 + 5 – 6	(9 – 10 ÷ 2) × 7	7 ÷ (4 – 3) × 2
8 + 7 × (4 – 3)	10 × (6 + 4 – 3)	8 × 10 + 6 – 2
6 + 3 – 4 ÷ 2	2 + 4 × 7 – 10	7 + 5 × 8 – 4
8 × 5 – 4 + 7	9 – 8 + 6 × 5	4 + 3 – 8 ÷ 2
4 × 6 ÷ (10 + 2)	3 × 8 – 2 + 7	5 + 9 ÷ 3 × 8
7 + 3 × 5 – 8	7 × (9 – 3 + 4)	6 + 10 × 7 – 8
(8 + 6 – 10) × 5	9 × 4 – 3 + 7	9 × 7 – 5 + 4
10 × 2 – 3 + 7	7 ÷ (2 × 8 – 9)	(10 – 7 + 4) × 2
4 × (9 – 8 + 3)	8 ÷ 2 – 3 + 6	(7 – 6 + 8) ÷ 9
10 + 5 × 4 – 6	3 + 9 × (6 – 5)	2 × (8 + 5 – 4)
6 × 8 ÷ 3 + 9	8 × 9 – 7 + 6	5 × (7 – 6 + 9)
7 × (10 + 2 – 3)	(7 – 3) × 2 + 5	5 + 8 × 6 – 2
4 × 7 – 10 + 9	10 + 9 × (8 – 7)	9 – 8 + 6 × 2
(8 ÷ 2 – 4) × 7	8 × 5 + 4 – 7	(8 ÷ 4 + 9) × 6
(7 + 3 – 4) × 2	(7 + 6 – 10) × 3	6 × 4 ÷ 8 + 9
7 + 10 × (5 – 3)	5 × 2 + 9 ÷ 3	2 + 8 × 3 ÷ 4
9 × (3 + 10 – 2)	4 × 7 – 3 + 9	3 × 10 + 8 – 7

$7 \times (2 + 4 - 5)$	$3 \times 10 + 8 - 6$	$9 \times 10 \div 3 + 6$
$5 \times 2 - 9 + 7$	$(6 + 2 - 4) \times 3$	$4 + 10 \times 9 \div 6$
$9 \times (10 + 5 - 7)$	$(7 + 3 - 4) \times 9$	$7 - 4 + 2 \times 9$
$(5 + 8 - 9) \times 2$	$(10 - 8) \times 4 + 5$	$(7 + 8 - 10) \times 2$
$(5 + 10 - 3) \div 6$	$10 \times 6 - 7 + 8$	$(10 \times 2) \div 4 + 6$
$8 \times 9 - 6 + 3$	$7 + 6 \times 4 - 8$	$2 + 6 \div (4 - 3)$
$(3 + 5) \times 8 - 7$	$10 - 4 + 8 \div 2$	$8 \div 2 \times 5 + 4$
$(9 + 6) \div 5 \times 8$	$(10 \div 2) \times 8 - 5$	$10 - 9 \div (2 + 7)$

APPENDIX D: DEMOGRAPHICS QUESTIONNAIRE

1. Sona ID number:
    - ❑ (Option to enter text)
2. Age:
    - ❑ (Option to enter text)
3. Biological Sex (i.e. sex you were assigned at birth):
    - ❑ Male
    - ❑ Female
    - ❑ Would rather not say
4. Gender with which you best identify yourself:
    - ❑ Male
    - ❑ Female
    - ❑ Other: (Option to enter text)
    - ❑ Would rather not say
5. How would you classify yourself?
    - ❑ Arab
    - ❑ Asian/Pacific Islander
    - ❑ Black
    - ❑ Caucasian/White
    - ❑ Hispanic/Latino
    - ❑ Multiracial
    - ❑ Other: (Option to enter text)
    - ❑ Would rather not say
6. Year in school:
    - ❑ Freshman
    - ❑ Sophomore
    - ❑ Junior
    - ❑ Senior
    - ❑ Graduate/Professional school
    - ❑ Other: (Option to enter text)

7. Major:
    - ❑ (Option to enter text)

8. Minor:
    - ❑ (Option to enter text)

9. Did you complete this study on a:
    - ❑ Desktop Computer
    - ❑ Laptop
    - ❑ Tablet (i.e. iPad)
    - ❑ Smartphone/Cellphone

APPENDIX E: INFORMED CONSENT

EXPLANATION OF RESEARCH

Title of Project: Memory of Words: A Categorization Task

Principal Investigator: Valerie Sims, PhD

Co-Investigator: Paulina Maxim

Faculty Supervisor: Valerie Sims, PhD

You are being invited to take part in a research study. Whether you take part is up to you.

- The purpose of this research is to observe differences in studying methods when learning and recalling words. The researchers wish to determine if categorizing words has a significant effect on memory recall. Implications involve finding better teaching methods that could be used to improve the encoding and retrieval of information.

- You will be randomly assigned to one of four conditions and will be shown several lists of words, one after the other. The words are common nouns in the English language and you will be asked to recall as many words as possible in a specific time frame. Following every word list, you will be asked to answer a few standard order-of-operations math problems to the best of your ability. You will not be penalized for any incorrect responses. Lastly, you will be asked to answer a brief demographics questionnaire. The entire study will be conducted online and you will not be asked to come into a lab.

- The expected duration of this study is approximately 1 Hour. There are no additional time commitments for this research.

You must be 18 years of age or older to take part in this research study.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints please contact Paulina Maxim, Undergraduate Student, Department of Psychology at paulinamaxim@knights.ucf.edu or contact Dr. Valerie Sims, Faculty Supervisor, Department of Psychology at valerie.sims@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been determined to be exempted from IRB review unless changes are made. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

APPENDIX F: IRB APPROVAL LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Determination of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Valerie K. Sims** and Co-PI: **Paulina Maxim**

Date: **July 12, 2018**

Dear Researcher:

On 07/12/2018, the IRB reviewed the following activity as minor modifications to human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Modification Type:	Study Title changed from "False Memory of Words: A Categorization Task" to "Memory of Words: A Categorization Task." Revised study Application, version 1.2, attached. Increased number of participants from 160 to 200 and total time of study from 30min to 1hr. Added questions to the demographic questionnaire and added math problems questionnaire. Revised 509 form, consent, and questionnaire uploaded in iRIS.
Project Title:	Memory of Words: A Categorization Task
Investigator:	Valerie K. Sims
IRB Number:	SBE-18-14088
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:

Kamille Chaparro

Signature applied by Kamille Chaparro on 07/12/2018 03:45:38 PM EDT

Designated Reviewer

REFERENCES

Brainerd, C., & Wright, R. (2005). Forward association, backward association, and the false-memory illusion. *Journal of experimental psychology: Learning, Memory, and Cognition, 31*(3), 554.

Cairncross, S., & Mannion, M. (2001). Interactive multimedia and learning: Realizing the benefits. *Innovations in education and teaching international, 38*(2), 156-164.

Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of experimental psychology, 71*(6), 858.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology, 58*(1), 17.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review, 4*(2), 271-276.

Gallo, D. A., & Roediger III, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469-497.

Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese-Roediger-McDermott paradigm. *American Journal of Psychology, 116*(1), 1-14.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67.

Kroes, W. H., & Libby Jr, W. L. (1971). Conceptual Encoding Categories in Free Recall. *Perceptual and Motor Skills, 32*(1), 267-269.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared.

McDermott, K. B. (1996). The Persistence of False Memories in List Recall. *Journal of Memory and Language, 35*(2), 212-230.

McDermott, K. B., & Roediger III, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*(3), 508-520.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition, 21*(4), 803.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-407.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494-500.

Russell, W. A., Jenkins, J. J., & Psychology, U. o. M. D. o. (1954). The Complete Minnesota Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test: Department of Psychology, University of Minnesota.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*(1), 49-60.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of experimental psychology, 70*(1), 122.

Underwood, B. J., & Richardson, J. (1955). Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*(1), 1.