THE ROLE OF MEMORY PERCEPTION IN PREDICTING PERFORMANCE AND FALSE MEMORY FORMATION

A Thesis By

CHRISTOPHER NGOV
ORICID iD: 0000-0002-9440-3710

California State University, Fullerton
Summer, 2021

In partial fulfillment of the degree:
Master of Arts, Psychological Research

Department:
Department of Psychology

Committee:
David Gerkens, Department of Psychology, Chair
Iris Blandon-Gitlin, Department of Psychology
Steven Smith, Department of Psychology

DOI:
10.5281/zenodo.5088975

Keywords:
false memory, metamemory, judgement of learning, memory perception, word lists

Abstract:
This study measured the predictive capabilities of metamemory on memory performance and false memory occurrences in associative and category word lists. Metamemory was assessed using Factors 1 and 2 from the Self-evaluation of Memory Systems Questionnaire (SMSQ) and judgements of learning using a betting paradigm. The results indicated a lack of relationship between how someone rated their memory and how they performed on a free recall task. However, the results indicated unique differences when analyzing the word lists separately. The inclusion of JOLs significantly increased the predictive capacity of the model for both types of word lists when the outcome variable was proportion correct of memory performance, however, was only significant for associative word lists when the outcome variable was false memory occurrences. These findings suggest possible differences between types of word lists in regard to false memory formation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I want to express the utmost gratitude for Dr. David Gerkens. His unconditional support and guidance throughout the entire process allowed me to be successful, and I am truly grateful for his mentorship, encouragement, and advice. I would also like to thank my committee members Dr. Steven Smith and Dr. Blandon-Gitlin for agreeing to serve on my committee and taking time to support me on this project.

I am also very grateful for my friends and other family members for encouraging and aiding me when I was stressed or flustered throughout the process. Thank you for your words of encouragement! Specifically, I would like to thank my both of my parents for their unwavering support on all of my aspirations and academic endeavors. Thank you for constantly believing in me, even when I did not believe in myself. Thank you for all the sacrifices you have made to provide me with a better life. I truly thank the both of you. Lastly, I want to thank my great grandmother for modeling kindness, patience, and perseverance.

CHAPTER 1

INTRODUCTION

Memory is a crucial construct as humans rely on it for both trivial and monumental forms of decision making. These can range from remembering to bring a coupon to the store to identifying the perpetrator from a lineup of suspects. Often times, however, our memories of previously encoded events may become altered or manipulated beyond our own awareness which may result in false or distorted memories. Humans experience an influx of stimuli daily which leads to encoding, storing, and retrieving endless amounts of information. As a result, false memories can become more frequent and occur more than we are aware of. Numerous studies have been performed that account for certain variables such as misinformation of events (Zhu et al., 2010) and age (Dennis et al., 2008) that may contribute to the formation of false memories. Many studies focus on manipulation of variables to create a false memory experience, while fewer studies have focused on individual differences in humans that may influence the likelihood of producing a false memory experience. The latter is important to study as individual differences may elucidate the extent to which specific qualities or beliefs may influence false memory recall in humans. An area that has been less explored in relation to false memory is the field of metacognition, which seeks to understand our awareness about our own cognitive or mental processes.

Metamemory and Memory Performance

Metamemory (MM) is a subcategory of metacognition which focuses on an individual's ability to understand their own judgements of memory, inclusion of memory strategies, and perception of confidence in regard to memory retrieval (Nelson, 2001). The assumption of the latter presumes that individuals have varying degrees of confidence which is based on their own subjective judgement about their own memory capability. The way an individual views their memory as a whole construct may also help guide the way they make decisions—overconfidence and under confidence in memory ability can influence the likelihood of overestimating or underestimating accuracy in task performance

(Leippe et al., 2006). This study investigates how an individual's perception of their own memories may influence the probability of producing false memories.

Judgements of learning (JOLs) is a popular metacognitive strategy used to subjectively assess the likelihood that currently presented information will be remembered in the future (Arbuckle & Cuddy, 1969; Rhodes, 2016). Studies that have employed JOLs typically expose participants to a cued-pair of words and have participants rate their confidence in predicting their ability to recall it later. Although there has been some research asserting that JOLs are not necessarily as impactful for future recall when compared to not making JOLs (Benjamin et al., 1998), other studies have argued that JOLs are effective for improving recall of future events (Doughtery et al., 2005; Zechmeister & Shaughnessy, 1980). This assumption is grounded in the cue utilization theory (Koriat, 1997) and the direct-access theory (King et al., 1980). Cue utilization posits that the presented items assume varying degrees of intrinsic (inherent properties of word), extrinsic (learning environment), and mnemonic cues (assessment to which the word has been learned) which affect the later recall of the word given the degree of such cues (Koriat, 1997; Doughtery et al., 2005). Similarly, the direct access theory assumes that each word has a "strength" which is the underlying basis of JOL ratings. Higher JOLs are correlated with this strength because the accessibility of the word affects the strength, ultimately affecting the JOL ratings. This is also demonstrated in trace access theory which focuses on the assessment of strengths based upon memory traces (Busey et al., 2000). Stronger items tend to be recalled because of their strength in recognition, thus accuracy and confidence covary in relation to memory trace strength. When monitoring contents of memory for encoded items, individuals heavily rely on the strength of which they remember the items and base their confidence judgements on this assessment (Busey et al., 2000).

Measures to Assess Memory Perception

Memory surveys and questionnaires assess an individual's perception of their cognitive ability. Popular scales such as the Multifactorial Metamemory Questionnaire (Troyer & Rich, 2002), Squire Subjective Memory Questionnaire (Squire et al., 1979), and Metamemory in Adulthood Questionnaire

(Dixon & Hultsch, 1988) help researchers study the effects of aging and metacognitive performance in older adults. These surveys consist of items that address memory loss or decline and assist through quick subjective assessments. Although these surveys and questionnaires have been considered quite reliable and valid, they primarily focus on older populations and utilize a more clinical approach that may not be suitable for all populations, especially younger adults. A newer survey that has been developed to address metamemory on different memory constructs is the Self-evaluation of Memory Systems Questionnaire (SMSQ; Tonkovic & Vranic, 2011). The SMSQ was designed to assess participant perception of different memory systems as well as memory perception via survey items. The questionnaire surveys different factors ranging from subjective perception of individual memory to more specific types of memory such as episodic or semantic memory.

While subjective assessment of memory provides valuable insight to researchers, how useful is it in predicting other factors such as recall accuracy? These subjective assessments guide basic decision-making processes but are also susceptible to false memory occurrences even for items that seem renown or obvious (Blake et al., 2015). Blake et al. (2015) found that individuals exuded extreme confidence in their ability to recognize and replicate the iconic Apple logo. Less than half of the participants were able to correctly recognize the logo when presented with inaccurate depictions of the correct logo and only 1 participant (out of 85) was able to correctly reconstruct the logo from memory. Their results demonstrate that high confidence judgments do not always lead to accurate objective performance. Previous research has had varied results with subjective memory predicting objective performance. Jopp and Hertzog (2007) found a relationship for subjective and objective memory. In their study, they measured memory belief to investigate if there was a relationship between how someone perceives their memory and cognitive performance. Cognitive performance was measured using a cued-pair recall task as well as other measures of perceptual speed, span, inductive reasoning, vocabulary, and comprehension. Interestingly, they used a regression and found that cognitive performance on different activities were partially mediated by memory perception. This demonstrates that memory perception does play a role in predicting performance on cognitive tasks, even when demographic variable such as age were controlled for. Those findings contrast with Zelinksi et al. (2001) who found no distinct relationship between the two. In their study they measured subjective memory using the Memory Functioning Questionnaire (MFQ) which uses different factors to assess general memory rating and forgetting. The objective memory task they used was a free recall task from a list of unrelated word items. The findings of this study suggest a small, if any, relationship between subjective memory assessments and objective memory performance. While the population was different between this study and that of Jopp and Hertzog (2007), the inconsistencies between the relationship between subjective and objective memory should continue to be investigated. Thus, more research needs to be done to identify any reliable relationships between the two variables.

Word Lists and False Memory

False memory and false recognition are widely studied because of their implication in everyday life and within more specific domains such as criminal proceedings. This has led researchers to create and study different ways to manipulate false memory experiences from within the laboratory. Some of the most common measures to elicit false recall is through the use of word lists. A popular measure used to assess false memory phenomena in a controlled laboratory setting is the Deese-Roediger-McDermott Paradigm (DRM; Roediger & McDermott, 1995). The DRM lists are also known as associative list due to the nature of the list items. Due to its simplicity and ease of administration, it has become one of the most well-known measures to assess false memory occurrences. Participants are presented with a list of words that are related to a critical lure word. The critical lure word is not presented, but participants often recall the word from the list even though it was never presented. One possible explanation for the DRM illusion arises from issues of reality monitoring which occurs when participants are unsure of whether the critical lure word was generated externally (actually presented) or if it is a product of internal generation. The robust responses on the DRM have allowed researchers to capture the essence of false memory in a laboratory setting.

Another type of list used to elicit false memories are categorized word lists (Smith et al., 2002). These lists are comprised of items that share a taxonomic category. For example, if the category was “vegetables,” the word list would consist of broccoli, lettuce, carrots, etc. To elicit false memories, these word lists tend to exclude highly dominant words that are semantically related to specific lists (Smith et al., 2002). For example, the word “orange” is highly associated with the category list of “fruit” and is often falsely recalled even when it is not presented (Smith et al., 2002). Because the dominant word tends to be highly associated with the list, it elicits falsely recalled memories when the words of that list are recalled. Critical lure words for category lists are usually selected based on the item that is most often recalled within the list.

Associative word lists rely on the list items to generate false memories. DRM word lists rely on backwards associative strength (BAS) which is the average tendency for list items to trigger or elicit a false memory during a recall or association test (Roediger et al., 2001). Recall and recognition of critical lure items heavily depends on the BAS—higher BAS increases the likelihood of recalling critical lure items (Knott et al., 2012). Another factor to predict performance on DRM tasks are the number of words recalled at test. Roediger et al. (2001), found an inverse relationship on DRM word lists between the number of words recalled at test and false recall of the critical lure. This means that the more words a participant recalls from the list, the less likely they are to demonstrate false recall for lure words. Critical lure words for categorized lists are selected by omitting the word with the highest recall frequency—this is referred to as output dominance. Unlike DRM lists, categorized word lists use output dominance to elicit false memories. To our knowledge, there is no research demonstrating a similar effect as seen in Roediger et al. (2001) using categorized word lists.

Research regarding these types of word list have found differing results that support different hypotheses regarding false memory formation. Associative word lists tend to be more in line with the Kirkpatrick hypothesis (Kirkpatrick, 1984; Smith et al., 2002). The Kirkpatrick hypothesis states that false memories are more prone during the encoding stage (Smith et al., 2002). Categorized list tend to follow the Deese hypothesis (Deese, 1959) which suggests that false memories are a product during the memory test. In a series of experiments, Smith et al. (2002) tested to see if categorized list may demonstrate indirect priming effects using stem completion tests. Their findings are contrasted with associative word lists which tended to be prone to indirect priming using stem completion tests, which suggests that associative words tend to be activated during encoding and may cause subsequent false memory errors (Smith et al., 2002). This is believed to be the case because during recollection, as items may be added that are similar to the concepts that are supposed to be recalled (Deese, 1959).

There has been, however, research arguing that both list types actually activate false memories during the study phase of the experiment. Dewhurst et al. (2009) examined both lists in an experiment and found that both DRM and categorized lists are only influenced by associations during the study period and not during the testing phase. The conflicting results from Smith et al. (2002) and Dewhurst et al., (2009) demonstrate a need for further investigation of list type to understand the time in which false memories are more susceptible.

The Proposed Study

The current study will investigate multiple questions relating to metamemory and false memory recall. The current study will investigate the relationship between the SMSQ (Factor 1 and Factor 2), sum of bets (JOLs), proportion of correct recall, and proportion of false recall. While survey responses and JOL ratings are both subjective to the individual, these types of reports remain one of the most crucial forms of data to tap into metacognitive processes (Jersakova et al., 2016). The basis of JOL ratings are partially rooted in a participant’s confidence of their memory ability (Serra et al., 2008). It is also important to recognize that JOL ratings and experiments tend to utilize more episodic memory, so understanding the relationship between JOLs and other types of memory would be insightful and contribute to knowledge of JOL decision making.

Hypothesis #1

It is hypothesized that sum of JOLs within the experiment and proportion of correct recall will positively correlate with the episodic memory factor on the SMSQ.

Hypothesis #2

Since JOL and confidence of memory perception have overlapping qualities, it is also hypothesized that the subjective experience factor on the SMSQ will correlate with sum of JOLs made during the experiment.

Hypothesis #3

Research on JOL accuracy has shown that under many circumstances, JOLs can provide fairly accurate predictions on future performance (Dunlosky & Nelson, 1994; Lovelace, 1984). A hierarchical regression can be analyzed with the SMSQ Factor 1 and Factor 2, then adding sum of bets second, with the outcome measure being the proportion of correct recall. This analysis would be run separately between the two types of word lists to see if there are any striking differences between category and associative word lists. This will aid in answering whether or not associative and category lists differ between metamemory and memory performance for correct recall. It is hypothesized that memory perception will be able to predict proportion correct on both types of word lists when entered into regression models.

Hypothesis #4

In line with the results of Roediger et al. (2001) and the information regarding JOL accuracy being quite reliable in predicting future performance, it is hypothesized that individuals who make higher JOLs would be able to correctly recall more words on DRM lists, which would possibly lead to less false recall of critical lure words. Alternatively, individuals who demonstrate lower JOLs may not recall as many words, which may possibly lead to more falsely recalled words on associative lists. As for categorized list, it is hypothesized that there may be a more positive relationship between number of correctly recalled words and falsely recalled critical lure words. Since categorized word lists rely more on output dominance and less on BAS and items specific activation, recall for correct list items and critical lure words may be positively correlated.

The use of JOLs as subjective memory, to our knowledge, has not been researched extensively with false memories through word lists. It is important to study this comparison because while everyone is susceptible to false memories, understanding which variables may lead to an increase in false memory recall is crucial. In this study, individual differences of subjective memory perception and its plausibility in predicting false memory recall will be investigated. Research on JOL accuracy has shown that under many circumstances, JOLs can provide fairly accurate predictions on future memory performance (Dunlosky & Nelson, 1994; Lovelace, 1984). If this assumption is true, participants will be fairly accurate in judging what they will and will not remember, but whether or not they will be susceptible to higher frequencies of false memories is unknown. Since JOL accuracy can provide accurate predictions on performance, this will lend insight as to whether or not DRM and category lists evoke false memories during encoding as stated in Dewhurst et al. (2009). Since JOLs are created during the encoding phase, JOLs should be better at predicting performance on both lists given the findings of Dewhurst et al. (2009). Alternatively, if JOLs only accurately predict performance on DRM word lists and not category word lists, then it would be more in line with the findings of Smith et al. (2002) which states that formation of false memory occurs at different times between the two list types.

CHAPTER 2

METHOD

Participants, Recruitment, and Procedures

All participants were undergraduate students attending California State University, Fullerton. Participants were recruited through the psychology department's research pool. Prior to data collection, four exclusion criteria were decided on to include participants in the final analyses. The exclusion criteria were as follows: scoring less than 50% on the math section, scoring three standard deviations from the mean on the experiment, completion of the entire experiment in less than 10 minutes, and taking 10 or more minutes on any of the given blocks within the experiment. Participants that violated one or more of the exclusion criteria were not included in the final analyses. The study was interested in adult cognition, so individuals must be 18 years or older to have participated. Participants also must be able to read and write in English. There were no exclusion criteria for participant demographics such as major, class level, gender, or race. A total of 132 participants completed the study. Twenty-one participants were removed from the final analyses for violating one or more of the stated exclusion criteria. A total of 111 participants were included in this study.

Participants will voluntarily signed up for time slots through the psychology department's research subject pool. All participants were shown an informed consent screen that details the necessary information required by the Institutional Review Board. Upon reading and agreeing to the informed consent, participants will be randomly assigned to complete either the experiment or survey first. Participants who were randomly assigned to complete the experiment first completed the survey portion after, whereas participants who were presented with the survey first were given the experiment afterwards. Participants were debriefed at the end of the experiment. The study will last approximately 30 minutes. Participants were awarded course credit for their applicable academic courses upon completion of the study.

Measures and Materials

The Self-evaluation of Memory Systems Questionnaire (SMSQ; Tonkovic & Vranic, 2011) is a survey that is comprised of 57 items. All test items are measured on a 5-point Likert-type scale ranging from “strongly disagree” to “strongly agree.” The items are divided into six different factors. The first factor is Subjective Evaluation which includes 13 items used to assess self-perception and contentment of memory (i.e., “I have a better memory than most of my peers.”). Factor 2 is Episodic Memory and is comprised of 14 items used to assess views on conscious experiences of remembering events (i.e., “I can always easily remember what I wore the day before.”). Factor 3 is Semantic Memory and is comprised of 15 items that are used to assess views on learning and remembering factual information (i.e., “It is enough for me to hear a word once and remember it.”). Factor 4 focuses on memory for numbers and is comprised of 6 items used to assess views on memory involving numerals or digits (i.e., “I can do calculations in my head relatively easily.”). Factor 5 focuses on visuospatial memory and is comprised of 5 items used to assess memory for facial, spatial, or locational memory processes (i.e., “When I am in a new city, I easily recall the places I have passed by.”). Factor 6 focuses on reminders and the extent to which they are utilized in everyday scenarios (i.e., “I often make to-do lists.”). It contains 4 items. While the SMSQ contains six factors, the analyses will primarily use Factors 1 and Factors 2.

The internal consistency for each Factor was calculated using Cronbach’s alpha. Factors 1 (subjective experience; 0.85), 2 (episodic memory; 0.83), 3 (semantic memory; 0.82), and 4 (memory for numbers; 0.81) were calculated and demonstrated relatively good alpha levels. Factors 5 (visuospatial; 0.74) and 6 (reminders; 0.78) demonstrated decent alpha levels. The overall internal consistency was estimated to be 0.92 for the SMSQ, demonstrating sufficient reliability.

Content validity was assessed by professional judgement from 6 experts in the field of cognitive science and or memory research. A few items were omitted due to an insufficient inter-rater reliability assessment and were removed from the test resulting in 57 total items in the survey.

Word Lists

DRM and category word lists were used to examine false memory performance. Three DRM word lists were taken from Roediger et al. (2001) which contained 15 items per list. Three category lists were retrieved from Barsalou (1985) which contained 15 words. Critical lure words for DRM lists were selected based on Roediger et al. (2001). Critical lure for category lists were determined by taking the word with the highest output dominance in the Barsalou (1985) taxonomic categories. Word lists are available in Appendix A.

Betting Paradigm

The betting task was administered through Qualtrics which is a website that allows users to create and collect survey data. This betting paradigm was adopted from Lin et al. (2019) but has been reconfigured to be administered on Qualtrics.

The betting paradigm consists of 6 blocks which contain 15 trials (each trial is one item on the word list). The blocks alternate between DRM and category word lists. At the beginning of each block, participants begin with the "betting" phase and are shown each word one at a time and are asked to make a "bet" (judgement of learning) on how likely they believe they will recall the word later (See Figure 1).

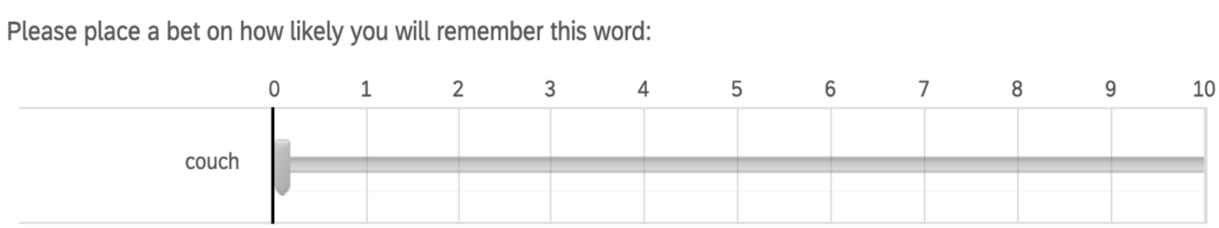

Figure 1. Participant view of the betting phase in the experiment for the word "couch."

Responses range from 0 (definitely will not remember) to 10 (definitely will remember) and participants use a sliding scale to make their bet. After all trials of a singular block have been viewed and rated, participants are presented with a arithmetic problem solving section where they must complete 10 math problems. Participants are instructed to be as quick and accurate as possible when completing this section. At the end of the block, participants are asked to recall the words that they remember rating at the start of the block. Participants use the keyboard to type in their answer and are given unlimited time to complete this section. Participants are instructed to complete a total of 6 blocks in the experiment.

Scoring

Points are awarded based on the bet made during the presentation phase. For example, if a participant places a bet of 10 points on the word “blanket” they will receive 10 points if they included it in the recall section. In the same scenario, if they do NOT recall the word, the participant will receive 0 points for that word. Total points that can be received for a singular block are based on the sum of the bets for the 15 words made by the participant for that block (totals range from 0 to 150). Performance on a block is calculated by taking the number of points received during the recall section and dividing it by the total possible points for the block.

Participants are asked to be as accurate as possible to the word that was presented during the presentation phase. Incorrect tenses (typing “walked” instead of “walk”) and plural version of the word (typing “mice” instead of “mouse”) will be considered incorrect responses. Incorrect spelling will be judged on how closely it resembles the presented list items.

Critical lure words that are recalled during the blocks as well as other words that did not appear in the presentation phase (that are not the intended critical lure word) were factored into as the proportion of false recall.

CHAPTER 3

RESULTS

A total of 111 participants were included in the study. Participants were randomly assigned to a condition that either presented them with the betting paradigm first followed with the SMSQ survey (Condition 1; 57 participants) or they were presented with the SMSQ survey first then given the betting paradigm (Condition 2; 54 participants). A between subject design assessed if there were any potential priming effects from administering the SMSQ survey first. To highlight any possible differences between the two groups, independent samples t-tests were used to compare Factor 1 and 2 of the SMSQ as well as total JOLs made during the experiment. For Factor 1 of the SMSQ, Condition 1 ($M = 38.67$, $SD = 4.86$) and Condition 2 ($M = 39.89$, $SD = 5.27$) showed no significant differences between groups, $t(109) = -1.27$, $p = .863$. For Factor 2 of the SMSQ, Condition 1 ($M = 46.17$, $SD = 6.97$) and Condition 2 ($M = 45.58$, $SD = 7.49$) showed no significant differences between groups, $t(109) = .427$, $p = .723$. For the total JOLs made throughout the experiment, Condition 1 ($M = 483.5$, $SD = 198.9$) and Condition 2 ($M = 470.98$, $SD = 166.74$) showed no significant differences between groups, $t(109) = .360$, $p = .094$. These t-test showed no significant differences between the two groups allowing the data to be collapsed into one comprehensive sample for further analysis.

Relationship between Sum of JOLs, Proportion Correct, and Factors on the SMSQ

A series of correlations were used to assess the relationship between the JOLs and SMSQ factors. Hypothesis #1 suggested that Factor 2 (episodic memory) of the SMSQ would correlate with the sum of JOLs from the betting paradigm and overall proportion of correct recall. Results from the correlation indicated that there was a significant positive correlation between Factor 2 and the sum of the JOLs from the betting paradigm; however, there was not significant association between Factor 2 and proportion correct.

Hypothesis #2 suggested that there would be a positive correlation when comparing Factor 1 (subjective evaluation) to the sum of JOLs from the betting paradigm. Results from the correlation show that there was a significant association between the two variables. These correlations demonstrate that individual subjective memory perception is associated with higher betting on judgement of learning recall tasks (See Table 1).

These correlations suggest that there is no association between how someone thinks about their memory and whether it affects their performance. Interestingly, there was a significantly positive correlation between sum of the JOLs made during the entirety of the experiment and the proportion correct from the entire experiment. This relationship demonstrates that JOLs may accurately predict future memory performance regardless of how an individual views their memory. See Table 1.

Table 1. Correlations Between Judgement of Learning Ratings, Proportion Correct, and SMSQ Factors

Variable	*M*	*SD*	1	2	3
1. SMSQ 1	39.3	5.09			
2. SMSQ 2	45.86	7.22	.574**		
3. SUM OF JOLs	477.07	182.4	**.197***	**.233***	
4. PROPORTION CORRECT	.65	.15	**.052**	**.184**	.306**

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed).

Predicting Proportion Correct on Associative Word Lists

A hierarchical linear regression was used to assess the predicative ability of memory perception on the proportion correct for associative word list items. The initial hypothesis suggested that sum of JOLs would be able to predict proportion correct on associative word lists. This provides insight as to whether or not sum of the JOLs made during the experiment can uniquely predicted performance on the free recall task. To test this, SMSQ1 and SMSQ2 were inputted in Model 1 and sum of JOLs was input in Model 2 with proportion correct being the outcome variable.

Model 1 included the SMSQ1 and SMSQ2 as predictive variables with proportion correct being the outcome variable. In Model 1, SMSQ 2 (B = -.097, p = .048) was found to be a significant predictor of proportion of correct recall while SMSQ1 (B = 0.231, p = .404) was not a significant predictor in Model 1 (See Table 3). As a whole, Model 1 was not significant when adding SMSQ1 and

SMSQ2, *F*(2,108) = 2.077, *p* = .13 and only accounted for 3.7% of the variance for the outcome variable (See Table 2).

Table 2. Hierarchical Regression for Associative Word Lists and Proportion Correct

	R	R^2	Adjusted R Square	SE	R Square Change	F Change	df1	df2	Sig
Model 1	.192	.037	.019	.167	.037	2.077	2	108	.13
SMSQ 1									
SMSQ 2									
Model 2	.376	.141	.117	.158	.104	13.013	1	107	.001
SMSQ 1									
SMSQ 2									
JOLs									

Table 3. Hierarchical Regression Coefficient Table for Associative Word Lists and Proportion Correct

	B	SE	Beta	t	Sig.	Zero-order	Partial	Part
Model 1	.502	.13		3.866	.0001			
SMSQ1	-.003	.004	-.097	0.838	.404	.036	-.08	.079
SMSQ2	.005	.003	.231	2.002	.048	.175	.189	.189
Model 2	.454	.124		3.67	.0001			
SMSQ1	-.004	.004	-.116	1.063	.29	.036	-.102	.095
SMSQ2	.004	.003	.173	1.567	.12	.175	.15	.14
JOLs	.001	.002	.331	3.607	.0003	.348	.329	.323

In Model 2, the sum of the JOLs for associative word list were entered to see if it would uniquely contribute to the predictive ability of the model. In Model 2, SMSQ1 (*B* = -.116, *p* = .29) and SMSQ 2 (*B* = .173, *p* = .12) were not significant. The addition of JOLs from associative word lists increased the model's predictive ability and accounted for 14.1% of the total variance. Model 2 was significant when the JOLs were added, *F*(3,107) = 5.88, *p* < .01.

The results of this analysis supports the hypothesis that using the JOLs from the experiment can increase the predictive ability of the model beyond what the participant generally thinks about their memory. This analysis further demonstrates that the way an individual thinks about their memory may be different or less important than other self-efficacious ratings or metamemory judgment measures.

Predicting Proportion Correct on Category Word Lists

A hierarchical linear regression was used to assess the predicative ability of memory perception on the proportion correct for category word list items. It was hypothesized that sum of JOLs would be able to predict proportion correct on category word lists. This provides insight as to whether or not sum of the JOLs made during the experiment uniquely predicted performance on the category word list recall task. To test this, SMSQ1 and SMSQ2 were input in a model 1 and sum of JOLs was input in model 2 with proportion correct being the outcome variable.

Model 1 included the SMSQ1 and SMSQ2 as predictive variables with proportion correct being the outcome variable. In Model 1, SMSQ 1 (B = -.051, p = .0.186) was not a significant predictor of proportion of correct recall. SMSQ 2 (B = 0.199, p = .088) was not a significant predictor in Model 1 (See Table 5). As a whole, Model 1 was not significant when adding SMSQ 1 and SMSQ 2, F(2,108) = 1.706, p = .186, and only accounted for 3.1% of the variance for the outcome variable (See Table 4).

Table 4. Hierarchical Regression for Category Word Lists and Proportion Correct

	R	R^2	Adjusted R Square	SE	R Square Change	F Change	df1	df2	Sig
Model 1	.175	.031	.013	.15	.031	1.706	2	108	.186
SMSQ 1									
SMSQ 2									
Model 2	.27	.073	.047	.147	.042	4.853	1	107	.03
SMSQ 1									
SMSQ 2									
JOLs									

In Model 2, the sum of the JOLs for category word list were entered to see if it would uniquely contribute to the predictive ability of the model. The addition of JOLs from category word lists increased the model's predictive ability and accounted for 7.3% of the total variance. Model 2 was significant when the JOLs were added, $F(3,107) = 2.79$, $p = .03$.

Table 5. Hierarchical Regression Coefficient Table for Category Word Lists

	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
Model 1	.540	.117		4.63	.0001			
SMSQ1	-.002	.003	-.051	-.438	.663	.064	-.042	-.04
SMSQ2	.004	.002	.199	1.721	.088	.17	.163	.163
Model 2	.521	.115		4.525	.0001			
SMSQ1	-.002	.003	-.078	-.679	.499	.064	-.065	-.06
SMSQ2	.003	.002	.162	1.412	.161	.17	.135	.131
JOLs	.0003	.0001	.213	2.203	.03	.235	.208	.205

Predicting False Memory on Associative Word Lists

It was hypothesized that higher JOL ratings would lead to less falsely recalled words for associative word lists. In order to test this, a series of correlations were used to identify if there was any relationship between SMSQ Factors, sum of JOL ratings, and false memory occurrences (See Table 6). The correlations revealed that there was a negative correlation between the sum of JOLs for associative lists and false memory occurrences. This supports the initial hypothesis that there would be a negative correlation between the two variables. This suggests that the higher the bets that an individual places on remembering words, the less likely they are to recall critical lure words on associative word lists. This is also similar to the results of Roediger et al. (2001) suggesting a negative correlation between correct recall of words and false memory occurrences.

A hierarchical linear regression was used to see if total JOLs for associative lists would be able to predict false memory frequency on a free recall task. In Model 1, SMSQ 1 and SMSQ 2 were entered as predictor variables with false memory occurrences as the outcome variable. This would provide insight as to whether or not sum of JOLs for associative lists made during the experiment would uniquely predict false memory occurrences on associative word lists.

Table 6. Correlations Between SMSQ Factors, False Memory Occurrences, and Sum of JOLs for Associative Word Lists

	1	2	3	4
1. SMSQ 1				
2. SMSQ 2	.574**			
3. SUM OF JOLs	0.16	.209*		
4. FALSE MEMORY OCCURRENCES	-.237*	-.083	**-.225***	
5. PROPORTION CORRECT	.036	.175	**.**348**	**-.315****

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

In Model 1, SMSQ 2 (B = .08, p = .486) was not a significant predictor of false memory occurrences. SMSQ 1 (B = -.283, p = .015) was significant in Model 1 (See Table 8). This means that according to Model 1, the episodic memory component was not a significant predictor of false memory occurrences while subjective evaluation of memory was found to be a significant predictor. As a whole, Model 1 was significant when adding SMSQ1 and SMSQ2, $F(2,108) = 3.469$, p = .035 and accounted for 6% of the total variance (See Table 7). These results suggest that associative word lists may rely more on subjective memory perception as opposed to an episodic memory component. These findings are inconsistent with van Bergen & Merckelbach (2009) who found that subjective memory assessment through a survey was not predictive of endorsing critical lure words on associative word lists.

Table 7. Hierarchical Regression for Associative Word Lists and False Memory

	R	R^2	Adjusted R Square	SE	R Square Change	F Change	df1	df2	Sig
Model 1	.246	.06	.043	.337	0.06	3.469	2	108	.035
SMSQ 1									
SMSQ 2									
Model 2	.317	.101	.076	.331	0.04	4.809	1	107	.03
SMSQ 1									
SMSQ 2									
JOLs									

Model 2 included the SMSQ 1 and SMSQ 2 in the first block with sum of JOLs added after to see if it could uniquely account for false memory occurrences. Model 2 was found to be significant overall, *F*(2,107) = 4.809, *p* = .03, which accounted for approximately 10% of the total variance. Similar to Model 2, SMSQ 1 was still significant while, SMSQ 2 still remained not significant. This further suggests that subjective evaluation of an individual's memory may be quite useful in determining whether someone would falsely recall a critical lure word, while also suggesting a user's rated episodic memory may not significantly be related to false memory recall.

Table 8. Hierarchical Regression Coefficient Table for Associative Word Lists and False Memory

	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
Model 1	.893	.261		3.41	.001			
SMSQ1	-.019	.008	-.283	-2.48	.015	-.237	-.232	-.231
SMSQ2	.004	.005	.08	.699	.486	-.083	.067	.065
Model 2	.953	.258		3.68	.003			
SMSQ1	-.018	.008	-.27	-2.41	.018	-.237	-.227	-.221
SMSQ2	.006	.005	.115	1.02	.31	-.083	.098	.094
JOLs	-.001	.0003	-.206	-2.19	.03	-.225	-.207	-.201

Predicting False Memory on Category Word Lists

For category list, it was hypothesized that there would be a positive correlation between the proportion correct made during the experiment and the number of false memory occurrences. The correlations revealed that there was a negative association between false memory occurrences on category word lists and proportion correct. The correlation did not support the initial hypothesis and was more in line with the findings of Roediger et al. (2001) which illustrated a negative correlation between correct recall and false recall on associative word lists. It is also important to note that sum of JOLs made during the experiment had no significant association between false memory occurrences. This is contrasted with the negative correlation that was found between sum of JOLs made during the experiment and number of false memory occurrences on associative word lists (See Table 9).

Table 9. Correlations Between SMSQ Factors, False Memory Occurrences, and Sum of JOLs for Category Word Lists

	1	2	3	4
1. SMSQ1				
2. SMSQ2	.574**			
3. SUM OF JOLs	.226*	.246**		
4. FALSE MEMORY OCCURRENCES	-.071	-.026	**-0.07**	
5. PROPORTION CORRECT	.064	.170	.235*	**-.246***

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

A hierarchical linear regression was used to examine if total JOLs for category lists would be able to predict false memory frequency on a free recall task using category word lists. In Model 1, SMSQ 1 (B = -.083, p = .479) and SMSQ 2 (B = -.022, p = .85) were not found to be significant and only accounted for less than 1% of the total variance (See Table 10).

Table 10. Hierarchical Regression for Category Word Lists and False Memory

	R	R^2	Adjusted R Square	SE	R Square Change	F Change	df1	df2	Sig
Model 1	.073	0.005	-0.013	0.29	0.005	0.288	2	108	0.75
SMSQ 1									
SMSQ 2									
Model 2	.093	0.01	-0.019	0.29	0.003	0.368	1	107	0.545
SMSQ 1									
SMSQ 2									
JOLs									

Both SMSQ Factors were not significant in predicting false memories on category word lists. In Model 2, the sum of JOLs were added to see if that would increase predictive capabilities of false memory performance on category word lists. The results for Model 2 demonstrated that the addition of JOLs did not produce significant results between the models. This means that the addition of sum of JOLs did not predict false memory occurrences and it only accounted for 1% of the total variance.

The analysis demonstrated that for category word lists, sum of JOLs made during the experiment did not have predictive value for false memory occurrences. This means that regardless of how someone rated their future performance using JOLs, it was not indicative on how likely they were to subsequently falsely recall lure words on category word lists.

Table 11. Hierarchical Regression Coefficient Table for Category Word Lists and False Memory

	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
Model 1	.328	.229		1.42	.156			
SMSQ1	-.005	.007	-.083	-.711	.479	-.071	-.068	-.068
SMSQ2	.001	.005	.022	.19	.85	-.026	.018	.018
Model 2	.339	.231		1.46	.145			
SMSQ1	-.004	.007	-.076	-.64	.524	-.071	-.062	-.062
SMSQ2	.001	.005	.033	.276	.783	-.026	.027	.027
JOLs	-.004	.0003	-.061	-.606	.545	-.07	-.059	-.058

The overall findings and differences between associative word lists and category word lists suggest possible differences between the list types. Associative word lists demonstrated that subjective assessment through a survey method as well as through JOLs were able to significantly predict false memory recall. These findings contrasts with category word lists in that both subjective memory assessment and inclusion of JOLs in a model was not enough to predict false memory occurrences in a hierarchical regression model.

CHAPTER 4

DISCUSSION

The first hypothesis was supported, and demonstrated that there was a relationship between both SMSQ Factors and sum of JOLs for the entire experiment. The results indicated that there is a positive relationship between how someone perceives their memory and how they decide to make bets on a betting paradigm. This means that an individual who rates themselves as having a better memory will more likely place higher bets on words to be remembered in a future recall task. The results show that higher ratings of subjective memory evaluation (SMSQ 1) is correlated with higher likelihood of believing they will remember the words later (JOLs). Similarly, episodic memory (SMSQ 2) was correlated with sum of JOLs during the betting paradigm. This suggests that the higher an individual rates their episodic memory, the more likely they will tend to place higher bets on words they will have to recall later.

While these findings suggest that there is a relationship between how someone thinks about their memory and how they think they will do, the results did not support the other hypothesis of a potential relationship between memory perception and memory performance across the experiment. While the results indicated that there was no relationship between how someone rated their memory on the SMSQ and how they actually performed, it did demonstrate that JOLs made during the experiment were positively correlated with overall memory performance. These findings were also similar to that seen in Saraiva et al. (2020), where they tested subjective assessment of memory ability in an eyewitness free recall experiment. In their experiment they wanted to test whether or not metamemory assessments (in a survey format) would have any predictive indication of recall in an eyewitness crime video. They also looked at how metamemory assessments may possibly predict general confidence and the degree to which an individual was confident. They noted that individuals who had higher self-rated memory were more likely to be accurate in relation to their confidence assessments. This means that while metamemory assessment did not demonstrate as strong of a predictive value as initially hypothesized, the metamemory assessment did prove to be helpful in predicting confidence to accuracy relationship for eyewitness free recall. In a meta-analysis, Beaudoin and Desrichard (2011) analyzed the literature to see if there was any relationship between memory self-efficacy (MSE) and performance. MSE can be defined as the subjective evaluation a person would make of their own memory when presented with a memory task. Interestingly, they found that memory performance was more correlated with the perceived ability to perform the specific task than it was with a general belief about an individual's memory. They also expressed that MSE and performance was greatly affected by the type of task—recall tasks tended to demonstrate stronger relationships than recognition based tasks. The overall findings from Beaudoin and Desrichard (2011) could explain why SMSQ 1 and SMSQ 2 did not demonstrate a relationship with task performance while JOLs ratings did. This is most likely because the SMSQ is measuring overall general belief about an individual's memory, while JOL predictions are used to uniquely assess future performance within the betting paradigm and specifically within each word list.

Similarly, Hertzog et al. (1990) investigated how memory self-efficacy may affect predicted performance on a free recall task. They used the Capacity portion of the Memory in Adulthood Questionnaire (MIA) and the Frequency of Forgetting aspect of the Memory Functioning Questionnaire (MFQ) to measure memory-self efficacy. The Capacity portion and Frequency of Forgetting variables of the MIA and MFQ have been shown to reliably measure MSE (Hertzog et al., 1990). The MIA uses questions that are presented on a Likert-type scale and instructs participants to rate their self-efficacy for statements such as "I am good at remembering names" (Hertzog et al., 1990). The MFQ Frequency of Forgetting portion is similar as it has participants judge how often or frequently they tend to forget occasions, things, or names. Participants in the study were asked to read the directions of the experiment, and then were asked how they thought they would perform on the free recall task. They found a positive relationship between how an individual rated their MSE and their initial prediction of performance on a free recall memory task. Notably, the researchers stated that this relationship was much stronger for MSE than it was for metamemory surveys alone, suggesting that predictions of performance are much more reliant on MSE than general metamemory.

This is possibly due to metamemory encompassing different types of memory, while MSE only focuses on the task at hand.

The second hypothesis focused on how perception of memory may be able to predict performance on a free recall task using two different types of word lists. For associative word list, the findings supported the hypothesis that performance would be predicted by metamemory measures such as JOLs. When analyzing associative word lists, the predictive ability of the model significantly increased when JOLs were added. This means that JOLs, or how someone thinks they will perform on a given task, may provide greater insight beyond how they just feel about their memory. For category word lists, it was also found that JOLs may be able to predict proportion of correct recall. While the inclusion of JOLs did lead to significance for category word lists, the extent to which the variance was uniquely accounted for was not as robust when compared to associative word lists.

Another question this study could help answer was the debate over false memory formation for different word lists. Dewhurst et al. (2009) compared both category and associative word lists in an attempt to explain the similarities between the onset of false memories. The researchers found that both lists evoked false memories during encoding and not during testing as other researchers have noted, particularly for category word lists (Smith et al., 2002). A critique following the results of Dewhurst et al. (2009) were the taxonomic categories used to formulate the lists. The category word lists used in that study were more representative of associative word lists, which does not provide evidence for the onset of false memories for category lists since category word lists rely on a participant's pre-existing taxonomic representations. In order to explore this, both associative and category word lists were employed to see if the results would be in favor of Smith et al. (2002) or Dewhurst et al. (2009). The results indicated that for associative lists, the predictive capabilities of the model increased as the total JOLs made during the experiment were considered. This finding suggests that JOLs do contribute to the predictiveness of false memory recall. Correlations also revealed a similar relationship to Roediger et al. (2001) regarding the relationship between number of correct recall and false memory occurrences.

The results also indicated that there was a negative correlation between ratings on the SMSQ 1 and proportion of false memory recall. Interestingly, these results differ from other published articles such as van Bergen et al. (2009). In their study, they utilized the Squire Subjective Memory Questionnaire (SSMQ) to measure subjective memory performance while also using DRM word lists to elicit false memories. While they found different results to the current study such as a positive correlation between subjective memory ratings and the number of correctly recall words, they did not find a significant association between scores on the SSMQ and recall of the critical lure word. The differences in presentation (visual vs. audio) may possibly account for the differing results of the study. Another possibility is that the entirety of the SSMQ is based on both subjective memory and episodic-like memory questions. The results of the current study indicated a positive correlation between SMSQ 1 (subjective evaluation), but not for SMSQ 2 (episodic memory). Given that the SSMQ incorporates both SMSQ factors into one composite score, this may lead to differences between results.

It was originally hypothesized that category word lists may demonstrate a positive relationship between correct recall and false memory recall. Interestingly, the results for category word list did not support the initial hypothesis and turned out to be more similar to associative word lists regarding the relationship between correct recall and false memory recall. This relationship was similar to the inverse relationship found in Roediger et al (2008) using associative word lists.

It is important to note that JOLs made for category word lists demonstrated little to no relationship with false recall. The findings revealed that JOLs were not predictive of false memory on category word lists. These findings suggest that there may be a difference between category word lists and associative word lists regarding false memory formation. Smith et al. (2002) posited that the onset of false recall for category lists may be more susceptible during recall instead of encoding. Since the current study sought to understand the predictive qualities of JOLs on false recall, the lack of relationship may provide evidence that the two word list types elicit false memories at different phases. Since JOLs are created during the encoding phase, they would more likely predict false recall on associative word lists which would be in line with Dewhurst et al. (2009) results on associative word lists. As a result, the findings seem to favor Smith et al. (2002) in that there may be a difference between word lists types and the processes involved in false memory recall.

Limitations

Due to the nature of the study and the online format, several limitations occurred that were unavoidable. While this study implemented tools to record the time it took to complete each question, block, and experiment, other factors such as undivided attention and multitasking were not able to be fully accounted for since it was not conducted in person under normal laboratory supervision. Other possible factors such as the lack of control of the environment the participant was in may lead to inaccurate results. In a usual laboratory setting, equipment and environment would be standardized across participants. In the experiment, participants were asked to complete the experiment while being undisrupted for a certain period of time as well as to remove any distractions from the area that may interfere with the experiment. Participants were reminded to use a computer and not to perform the experiment on a tablet or mobile device. While participants were reminded of these restrictions, it is unknown to the extent to which they were followed which may ultimately alter performance between individuals.

Another factor that was not accounted for was the motivation or effort of the participants. Cognitive tasks generally require participants to perform to the best of their ability to accurately measure the desired cognitive construct. Since these tasks were administered asynchronously with no supervision, effort and motivation were more difficult to assess than if compared to an in-person study. A series of questions that asked participants how much effort or motivation they experienced while performing the task would have provided valuable insight that may illuminate any possible discrepancies between metamemory ratings and performance.

Future Directions

In order to better account for these limitations, a replication can be performed in-person to see if the findings would hold up. This would provide insight as to whether or not the online distribution method of a cognitive task would replicate the findings from this study. Due to the repetitiveness of the experiment, motivation may play a role in how willing a participant will perform at their best. A lack of motivation to perform the task may ultimately lead to lower ratings of perceived performance (JOLs) which may lead to inaccurate results. With this is mind, it may be important to include or assess motivation levels before and after the experiment to see if it may influence performance during the experiment.

APPENDIX A

WORD LISTS IN EXPERIMENT

Associative Word Lists *(bolded items are the critical lure words for that list):*

smell (BAS: .29)	***music (BAS: .22)***	***doctor (BAS: .24)***
Nose	Note	Nurse
Breathe	Sound	Sick
Sniff	Piano	Lawyer
Aroma	Sing	Medicine
Hear	Radio	Health
See	Band	Hospital
Nostril	Melody	Dentist
Whiff	Horn	Physician
Scent	Concert	Ill
Reek	Instrument	Patient
Stench	Symphony	Office
Fragrance	Jazz	Stethoscope
Perfume	Orchestra	Surgeon
Salts	Art	Clinic
Rose	Rhythm	Cure

Category Word lists (*bolded items are the critical lure word for that list*):

chair	***pants***	***apple***
couch	socks	kiwi
table	shoes	berry
desk	bra	pear
bed	skirt	raspberry
dresser	jacket	plum
lamp	sweatshirt	strawberry
sofa	hat	banana
cabinet	tie	peach
television	pantyhose	grape
stool	slacks	orange
rug	shirt	tangerine
refrigerator	dress	blueberry
stereo	shorts	lemon
bedstand	blouse	cherry
shelf	coat	lime

APPENDIX B

SMSQ SURVEY

All questions were presented on 5-point Likert-type scale ranging from “strongly disagree” to “strongly agree.”

Factor 1 (Subjective Evaluation)

1. I believe I recall the details of certain events better than others do.
2. I always remember to bring things I have promised.
3. Generally speaking, I am forgetful.
4. It often happens that I come to the store and can't remember all the things I wanted to buy.
5. I have noticed that I have better memory than others.
6. I wish I had better memory.
7. I have better memory than most of my peers.
8. I am always able to memorize things I really want to remember.
9. I can memorize large amounts of information in a short time period.
10. My memory is as good as it was 5 years ago.
11. People who know me believe I have a good memory.
12. When disputable, I rely on my own memory more often than somebody else's memory.
13. It rarely happens to me that a name or a word is 'on the tip of my tongue' and that I can't remember it.

Factor 2 (Episodic Memory)

1. I believe that memory is important.
2. I immediately memorize names of the people I get acquainted with.
3. I remember most of the names and surnames of my elementary school classmates.
4. I remember the dates of important events in my life.

5. I can always easily remember what I wore the day before.
6. I can easily remember where I have spent New Year's Eve in the last 5 years.
7. I remember where something is when window shopping.
8. I know the birthday dates of most of my friends.
9. I can easily associate a face to a name.
10. I remember the name and the face of my first elementary school teacher.
11. I rarely forget the time and the place where I am to meet someone.
12. I do not forget to congratulate birthdays to people who are close to me.
13. I can easily remember what I was doing on a certain day last week.
14. I correctly remember important addresses (friends, work, university, parents, etc.).

Factor 3 (Semantic Memory)

1. It is enough for me to hear a new word once and remember it.
2. In school, it was never a problem for me to learn a poem by heart.
3. I can easily memorize lyrics.
4. I can easily remember the chorus of a song I just heard.
5. If I read the morning newspaper, I can remember the headlines later in the afternoon.
6. I have difficulties in remembering the names of the characters in the book I recently read.
7. I remember the names of the authors of the books I've read.
8. I remember the stories I have heard in my childhood.
9. I have a good recollection of the plots in the movies I have seen.
10. When I watch a TV-show, I can always remember what happened in the previous episode.
11. Even if I forget to bookmark the page of the book I am reading, I can easily find the page I was on.
12. When I watch a movie, I easily remember even the less important characters.
13. I have a good memory for jokes.

14. I know a lot of quotations and sayings.
15. When I watch a movie, I can easily recall other movies in which I have seen the same actors.

Factor 4 (Memory for Numbers)

1. I can say a seven-digit number backwards when I hear it.
2. I can do calculations 'in my head' relatively easily.
3. I can easily memorize the time schedule of a bus or train.
4. I can easily memorize a new telephone number without writing it down.
5. I know almost all the important telephone numbers by heart.
6. I have a good memory for numbers.

Factor 5 (Visuospatial Memory)

1. When I am in a new city, I can easily recall the places I have passed by.
2. I remember the faces of the people when I first meet them.
3. I easily recognize pictures I have seen before.
4. I have a good recollection for people's faces.
5. I can return to the route that I have passed through only once, even after a long period of time.

Factor 6 (Reminders)

1. I often make 'to-do' lists.
2. I use reminders for everyday activities on a regular basis.
3. If I have to buy more than five things, I make a list.
4. I often use some form of alarm as a reminder of an activity in the near future.

REFERENCES

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55.

Blake, A. B., Nazarian, M., & Castel, A. D. (2015). Rapid Communication: The Apple of the mind's eye: Everyday attention, metamemory, and reconstructive memory for the Apple logo. https://doi.org/10.1080/17470218.2014.1002798

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: learning, memory, and cognition*, *11*(4), 629.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*(1), 26-48. https://doi.org/10.3758/BF03210724

Dennis, N. A., Kim, H., & Cabeza, R. (2008). Age-related differences in brain activity during true and false memory retrieval. *Journal of Cognitive Neuroscience*, *20*(8), 1390-1402. https://doi.org/10.1162/jocn.2008.20096

Dewhurst, S., Bould, E., Knott, L. and Thorley, C. (2009). The roles of encoding and retrieval processes in associative and categorical memory illusions. *Journal of Memory & Language, 60*(1), 154-164. https://doi.org/10.1016/j.jml.2008.09.002

Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1988). The metamemory in adulthood (MIA) questionnaire. *Psychopharmacology Bulletin*, *24*(4), 671-688.

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition*, *33*(6), 1096-1115.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565. http://dx.doi.org/10.1006/jmla.1994.1026

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*(2), 215.

Jersakova, R., Moulin, C. J., & O'connor, A. R. (2016). Investigating the role of assessment method on reports of déjà vu and tip-of-the-tongue states during standard recognition tests. *PLOS One*, *11*(4), e0154334.

Jopp, D., & Hertzog, C. (2007). Activities, self-referent memory beliefs, and cognitive performance: evidence for direct and mediated relations. *Psychology and Aging*, *22*(4), 811.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology*, 329-343.

Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychological Review*, *1*(6), 602.

Knott, L. M., Dewhurst, S. A., & Howe, M. L. (2012). What factors underlie associative and categorical memory illusions? The roles of backward associative strength and interitem connectivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 229.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349.

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Stambush, M. A. (2006). Effects of social-comparative memory feedback on eyewitnesses' identification confidence, suggestibility, and retrospective memory reports. *Basic and Applied Social Psychology*, *28*(3), 201-220. https://doi.org/10.1207/s15324834basp2803_1

Lin, G. C., Smith, R. N., Parlett-Pelleriti, C., Jones, M. R., & Jaeggi, S. M. (2019). I bet I'll remember "biochemistry"–Meta-memory as a function of lexical features and language background. https://doi.org/10.7939/r3-95m7-9p48

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766. http://dx.doi.org/10.1037/0278-7393.10.4 .756

Nelson, T.O. (2001). Metamemory, Psychology of. *International Encyclopedia of the Social & Behavioral Sciences, 26,* 9733-9738. https://doi.org/978-0-08-043076-8

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. K. Tauber (Eds.), Oxford handbook of metacognition (pp. 65–80). Oxford University Press.

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. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407 (2001). https://doi.org/10.3758/BF03196177

Saraiva, R. B., Hope, L., Horselenberg, R., Ost, J., Sauer, J. D., & van Koppen, P. J. (2020). Using metamemory measures and memory tests to estimate eyewitness free recall performance. *Memory*, *28*(1), 94-106.

Serra, M. J., Dunlosky, J., & Hertzog, C. (2008). Do older adults show less confidence in their monitoring of learning?. *Experimental Aging Research*, *34*(4), 379-391.

Smith, S. M., Gerkens, D. R., Pierce, B. H., & Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: The Kirkpatrick and Deese hypotheses. *Journal of Memory and Language*, *47*(3), 436-447. https://doi.org/10.1016/S0749-596X(02)00012-8

Squire, L. R., Wetzel, C. D., & Slater, P. C. (1979). Memory complaint after electroconvulsivetherapy: Assessment with a new self-rating instrument. *Biological Psychiatry, 14*, 791–801.

Tonković, M., & Vranić, A. (2011). Self-evaluation of memory systems: Development of the questionnaire. *Aging & Mental Health*, *15*(7), 830-837.

Troyer, A. K., & Rich, J. B. (2018*). Multifactorial memory questionnaire: Professional manual*. Baycrest Centre for Geriatric Care.

Van Bergen, S., Jelicic, M., & Merckelbach, H. (2009). Are subjective memory problems related to suggestibility, compliance, false memories, and objective memory performance?. *The American Journal of Psychology*, 249-257.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*(1), 41-44.

Zelinski, E. M., Burnight, K. P., & Lane, C. J. (2001). The relationship between subjective and objective memory in the oldest old: Comparisons of findings from a representative and a convenience sample. *Journal of Aging and Health*, *13*(2), 248-266.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... & Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, *18*(5), 543-555. https://doi.org/10.1080/09658211.2010.487051