# The Effect of Metacognitive Judgment and Spaced Rehearsal on Map Learning

An Honors Thesis for the Department of Psychology

Kishandra Anne Patron

Author Note

Holly A. Taylor, Ph. D, Department of Psychology, Tufts University;

Lauren A. Mason, Department of Psychology, Tufts University;

Ayanna K. Thomas, Ph. D, Department of Psychology, Tufts University.

Tufts University, 2021

The Effect of Metacognitive Judgment and Spaced Rehearsal on Map Learning

## Abstract

Map learning comprises the use of bird's-eye-view representations of an environment to encode spatial information, such as the locations and distances between landmarks. In the present study, I used a Judgment of Learning (JOL) methodology to examine the role of explicit metacognitive monitoring and study repetition in map learning. Researchers investigate the role of metacognitive monitoring in learning by measuring participants' JOLs, which are predictions of future success at recalling learned information. Previous work using word pairs argued that producing JOLs may affect control processes and enhance memory by influencing people's learning goals, a phenomenon called *JOL-reactivity*. In my experiment, participants studied one block of landmark pairs in a verbal context and another block in a spatial context. In each block, participants studied half the landmark pairs one time and half the landmark pairs three times. Following study, participants either produced delayed JOLs to studied items or completed delayed random number generation (RNG), also in context of the studied items. Participants lastly tested their memory in a non-spatial context. We found that participants recalled stimuli more accurately when study and test spatial contexts aligned. Moreover, participants produced more accurate JOLs when they studied stimuli three times in the same spatial context as test. My experiment offers preliminary and exploratory insight into how metacognitive monitoring and restudy impact map learning. Through greater understanding of the relationships between metacognition and map learning, we can better monitor our learning of new environments to shape more accurate map recollection during future test.

## Introduction

When planning for a road trip, there are several options to learn a designated route. You can print out and annotate a MapQuest planner in advance, taking notes on all the surrounding landmarks as depicted on the map and (re)rehearsing the route once a day leading up to your date of departure. On the other hand, you can input your destination into your smartphone GPS on the day you leave, studying your route once and casually observing spatial relationships in your mobile map. Both strategies exemplify map learning, which is the use of allocentric visualizations of an environment to encode spatial information such as the locations and distances between landmarks (Coluccia, 2007). Meanwhile, they differ in their degree of metacognitive processing, which is the interplay between monitoring one's own thoughts and exercising control to achieve a desired learning goal (Nelson & Narens, 1990). The MapQuest planner method requires more active goal-directed study of the map, whereas the day-of GPS method implicates less significant self-regulated study during map observation. They also differ in their degree of study repetition: whereas the MapQuest planner method requires repeated study of the route, the day-of GPS method only requires one study of the route.

Though employing metacognition and restudy strategies may potentially improve performance outcomes in map learning, minimal research has focused on the role of metacognition and restudy in learning through maps. The present study investigates how metacognitive monitoring and study repetition influence memory monitoring and recall accuracy during map learning.

## Map Learning

Maps provide users an efficient means of learning an environment. When studying a map of a spatial environment, observers can readily access configural information from the survey perspective, which is characterized by a bird's-eye viewpoint (Siegel, Kirasic, & Kail, 1978; Thorndyke & Hayes-Roth, 1982; Pazzaglia & Taylor, 2007; Taylor, Naylor, & Chechile, 1999). Yet, maps are complex stimuli, as they concurrently exhibit spatial and verbal information corresponding to an organized network of features. Given that learners have access to a variety of different details when studying a map, learners benefit from shifting between studying landmarks' relative and absolute locations. Landmarks' relative locations are their positions comparative to other nearby landmarks, whereas their absolute locations are their fixed positions, defined with respect to a structured system of coordinates (Allen et al., 1996; Coluccia, 2007).

### *The Role of Spatial Context in Learning and Memory*

When learning new information, the spatial context in which the information appears influences how and what one learns. Whereas studying information in a verbal context guides how one encodes new material, examining information in both verbal and spatial context facilitates learning and memory in a different way. The Conjoint Retention Hypothesis states that participants recall more text information when using a topographical map in conjunction with their readings (Verdi & Kulhavy, 2002). With a spatial organization supplementing the studied information, participants may more readily extract cross-code connections between the verbal and spatial information, a skill also known as dual coding (Clark & Paivio, 1991).

When viewing a map, users examine stimuli in different formats, namely as feature and structural information. Feature information constitutes individual icon and word labels that convey and facilitate memory encoding and retrieval. Meanwhile, structural information consists of the spatial network where discrete respective map features are embedded with fixed distances from each other (Kulhavy et al., 1992). Researchers claim that such structural internal markings allow for the “hooking” of feature information, thus increasing the efficacy of encoding structural information (Stamm et al., 1999). In my study, I am interested in how participants’ metacognitive processes may influence how they encode map feature and structural information to better recall associated text information. The act of monitoring their learning may influence how much learners can “hook” onto and remember information from maps.

### ***The Role of Spatial Skills in Map Learning***

In addition to learning goals, individual differences in spatial cognitive skills play a role in map learning. Researchers claimed that visuospatial working memory plays an essential role in map learning, as demonstrated by the empirical observation that a concurrent spatial tapping task impairs landmark positioning (Coluccia, 2007). Given these results, we turn our attention to how spatial cognitive skills moderate map learning.

Spatial cognitive skills comprise the ability of visualizing, in a given environment, one’s current orientation or an orientation differing from their own.

Similarly, spatial cognition recruits one's competence at updating their location in space as a result of self-motion (Hegarty et al., 2002). These skills play a part in map learning, for through spatial orienting and updating, one may effectively encode and reconstruct their imagined position with respect to the map environment. The Santa Barbara Sense of Direction Scale (SBSOD) is a self-report measure that correlates strongly with spatial updating and acquisition of spatial knowledge (Hegarty et al., 2002). I am utilizing the SBSOD in my study to observe how participants' sense of direction influences their memory performance in the map context.

## Metacognitive Monitoring

Reflecting on and remodeling study methods offers an efficient means of promoting productive learning. According to Nelson and Narens (1990), metacognition consists of two parts: (i) *monitoring,* which is the self-assessment of learning success or the future retrievability of content, and (ii) *control*, which is the regulation of learning through the employment of certain strategies. Researchers examine the role of metacognitive monitoring in various learning tasks such as multimedia learning, motor learning, and facial processing (Eitel, 2015; Rhodes, Simon & Bjork, 2001; Sitzman, & Rowland, 2013). However, the majority of work on metacognitive judgments of learning is confined to verbal materials, namely paired associate word-pairs (Rhodes, 2016).

One way experimenters measure participants' metacognitive monitoring processes is through prospective judgments of learning (JOLs), where participants make predictions about the future retrievability of material during the learning phase. Such judgments may be implemented experimentally through a continuous percentage scale (i.e., 0-100%) or a discrete rating scale (i.e., 1-7). Participants may produce JOLs immediately after studying an item or following a delay of intervening items (Double, Birney, & Walker, 2018).

Participants show reactivity to JOLs when those who judged their learning recall a significantly different amount of information from those who did not judge their learning. Metacognitive studies with word-pair stimuli have asserted that subjects who produce JOLs during study demonstrate positive JOL reactivity, characterized by increases in their retention and retrieval of studied materials during test, compared to their peers in the no-JOL control condition (Janes, Rivers, & Dunlosky, 2018). Cognitive scientists have hypothesized that the act of producing a JOL may affect how individuals attend to information, thus improving their monitoring processes and decreasing the likelihood that they engage in ineffective learning practices (Mitchum et al., 2016; Double et al., 2018).

In my study, I strive to extend these findings to more complex learning materials. Specifically, I am interested in how JOL production may differentially affect how one learns landmark pairs in verbal or spatial contexts.

### *The Role of JOL Context in Memory Performance*

We have witnessed how metacognition positively influences memory; now, it is important for us to understand what study settings are most conducive to JOL advantages. Memory researchers aim to define what learning contexts optimize the benefits of JOL production on test-related memory performance.

They study *JOL accuracy,* which is a learner's success at predicting their likelihood of accurately recalling a studied item (Dunlosky & Nelson, 1992; Rhodes & Tauber, 2011; Janes, Rivers, & Dunlosky, 2018). In other words, researchers investigate how learners' predictions of future memory performance correlate with actual memory performance, for example how items assigned a higher JOL (e.g., of 90%) may correlate with a recall accuracy distinct from items assigned a lower JOL (e.g., of 50%). Learners' predictions of future memory performance do not always correlate with actual memory performance, which thus motivates researchers to better understand what JOL contexts best improve predictive accuracy and memory performance.

**JOL Presentation.** JOLs may have different formats, and researchers have explored which format leads to the greatest JOL accuracy. In a mixed design JOL reactivity experiment using word-pair stimuli, where participants either made a JOL while viewing (a) the cue word of each word-pair or (b) the complete cue-target word-pair, researchers defended that JOL accuracy is significantly greater for participants who view only the cue word of each word-pair during JOL production (Dunlosky & Nelson, 1992). In my map learning experimental design, I implement JOLs in a cue-only format, and to further promote my choice, I next highlight significant hypotheses regarding the impact of study format on later test performance.

**Transfer Appropriate Monitoring (TAM) Hypothesis.** The study context in which one produces a JOL likely impacts their accuracy at predicting their future success at recalling learned information. As described by the Transfer

Appropriate Monitoring (TAM) Hypothesis, monitoring accuracy is dependent on how closely the JOL context mirrors the processes at test (Dunlosky et al., 2005). The congruency of study and test contexts impacts predictive accuracy. Specifically, greater predictive accuracy is seen when the studied stimulus is similar or identical to the cue at test. Conversely, a decrement in predictive accuracy is observed when the studied stimulus differs from the cue at test, or if the information at study requires processing that is not relevant for the test. Accordingly, in Nelson & Dunlosky's aforementioned study (1992), because the cue-word-only JOL context more closely simulates the test than the cue-target word-pair JOL context, JOL accuracy is significantly greater for participants who view only the cue word of each word-pair during JOL production.

**Transfer Appropriate Processing (TAP) Hypothesis**. Complementary to the Transfer Appropriate Monitoring (TAM) hypothesis is the Transfer Appropriate Processing (TAP) hypothesis. Morris et al. (1977) hypothesizes through the TAP hypothesis that the congruency of cognitive processes recruited during learning and test contexts predicts later test performance. According to the TAP hypothesis, researchers can predict that participants' performance differs in paradigms where study context misaligns from test context, compared to participants' performance in paradigms where study and test context align.

In my experiment, participants are studying and producing JOLs for stimuli in either a spatial or non-spatial context, but their memory is always tested in a non-spatial context. I examine how the misalignment of cognitive processes used for stimuli with contrasting study and test contexts may influence participants' memory performance.

### ***The Role of JOL Timing in Memory Performance***

JOL timing is an important component of JOL study context: participants may produce a JOL immediately after studying Item A, or they may switch to a studying a fixed set of items before revisiting and producing a JOL for Item A. In their within-subject paired-associate experimental design, where participants either produced a JOL (a) immediately after viewing each word-pair or (b) after a 10-item delay, Nelson and Dunlosky (1991) showed the *delayed JOL effect*, such that relative JOL accuracy is significantly greater for items judged after a delay.

Later research generally replicated the delayed JOL effect (Kelemen & Weaver, 1997; Koriat & Ma'ayan, 2005). Rhodes and Tauber (2011) demonstrated through their meta-analysis of 45 delayed JOL effect studies that delayed JOLs were marked by approximately one standard deviation ($g = 0.93$) increase in gamma correlations compared with immediate JOLs. Their finding shows support that, participants are more likely to remember studied items rated with high delayed JOLs and less likely to remember studied items rated with low delayed JOLs, with greater consistency than immediate JOLs.

My study utilizes delayed JOLs. To provide more confidence for my design, I next summarize a framework explaining the utility of delayed JOLs when studying metamemory.

**Monitoring Dual-Memories (MDM) Principle.** The timing of JOLs likely impacts whether people are using short-term or long-term memory to judge their learning. According to the Monitoring Dual-Memories (MDM) Principle, during learning, participants monitor both their short-term and long-term memory. When participants make JOLs immediately after study, information from short-term memory about the to-be-judged stimulus pair is interfering with information from long-term memory, thus hindering the accuracy of their immediate JOLs (Nelson & Dunlosky, 1991). Meanwhile, delayed JOLs are less likely to withstand noise from short-term memory and are more likely to accurately reflect long-term memory information, thus demonstrating more robust memory performance benefits.

Recent functional brain imaging studies provide evidence supporting the MDM principle. In particular, Kelley et al. (2020) used functional MRI to measure participants' brain activity as they studied and made JOLs for paired associates either (a) immediately following presentation or (b) after a delay. Consistent with Nelson and Dunlosky's MDM Principle (1991) that people retrieve information from long-term memory to inform delayed JOLs, Kelly and colleagues' neuroimaging results showed that activation patterns associated with past experience are more strongly linked with delayed JOLs than immediate JOLs.

### *The Influence of Repeated Study JOLs on Memory Performance*

Though we have observed how repeated study improves memory, less is known about how repeated study and metacognitive monitoring interact. Metacognition researchers have turned their attention to examine how producing JOLs during repeated study may influence memory performance during later test.

In particular, researchers have investigated how repeated study influences JOL confidence and influences predictions of future retrieval success. Chen, Zhang, and Liu (2019) demonstrated that when judging items, participants associated items they retrieve more fluently (i.e., with greater confidence and faster reaction time) with higher JOLs. Their strategy yielded higher correlation between JOL ratings and memory accuracy in final test trials compared with initial restudy trials, which suggests that participants readily monitor their memory processing and refer to experiences from their retrieval practice when assessing and predicting their learning progress. I am interested in what shapes effective metacognitive monitoring, for through accurate monitoring of our memory, we can improve how we regulate and support our memory encoding and retrieval processes.

Meanwhile, other researchers have stressed the difference in recall accuracy between judged, spaced rehearsal items (i.e., study repetitions of Item A are intervened by a finite set of Item X study repetitions) and judged, mass rehearsal items (i.e., study repetitions of Item A are successive). Logan, Castel, and Viehman (2012) demonstrated how participants recall significantly more judged, spaced rehearsal items than judged, mass rehearsal items. Participants recalled the fewest judged items that were studied only once. Similarly, researchers have observed significantly increased JOL accuracy for spaced restudy compared to massed restudy: as lag increased, final recall performance increased as JOL ratings decreased, suggesting JOL sensitivity to spaced restudy (Pyc & Rawson, 2012).

In my study, I strive to determine how making delayed JOLs during map learning influences memory performance. My experiment builds onto research efforts to characterize studies of reactivity in contexts beyond word-pair settings. I am examining the role of restudy in map learning by assigning half my experiment stimuli to be presented once and the other half to be presented three times in a spaced rehearsal design, as previous research has affirmed its utility over massed study repetition.

## Bridging the Two Cognitive Domains

Together, map learning and metacognitive monitoring require greater working memory resources than either of the tasks alone. Maps are complex stimuli that require attention to both the verbal and spatial features belonging to a structured system of coordinates, meanwhile metacognitive monitoring commands active rehearsal of stored memory to predict likelihood of successful future recall. Because maps are more complex, they recruit greater working memory resources, which translates into a higher cognitive load. Moreover, the act of producing a JOL during memory retrieval also commands more cognitive resources than passively studying information, thus inducing a greater cognitive load. Researchers have claimed that metacognitive monitoring while using a navigation display incurs one of the following cognitive loads: intrinsic, which results from processing information necessary for task; extraneous, which results from processing irrelevant features; and germane, which results from relevant and pertinent cognitive activities necessary for successful performance (Sweller, van Merrienboer, & Paas, 1998). Worded differently, germane load enhances a learner's ability to retain information meaningfully (Paas, Renkl, & Sweller, 2003).

Given that metacognitive monitoring has yielded memory performance advantages in word-pair contexts, it would be reasonable to predict metacognitive monitoring benefits in non-word-pair contexts. Nilsson and Mayer (2002) found that when utilizing a map to navigate a website's format, learners produce incomplete cognitive models because they can only attend to limited sections of the website. This is likely due to extraneous cognitive load. Researchers sought to understand how metacognitive monitoring may incur germane cognitive load to control more effective information processing. Building on previous research, Scott and Schwartz (2007) utilized a hypermedia environment-based experiment to demonstrate how, when exposed to a complex map, learners with high metacognitive skill incur high germane cognitive load and perform better in structured- and free-recall tasks, contrary to learners with low metacognitive skill who incur high extraneous cognitive load. The performance differences suggest that learners with better metacognitive skills extracted deeper comprehension of the website using their abilities to critique and evaluate the potential applications of studied information.

Although previous research has reported preliminary findings on metacognition's benefits on map learning (Nilsson & Mayer, 2002; Scott & Schwartz, 2007), it is critical that metamemory researchers push efforts further. Spatial researchers have demonstrated how navigational aids divide attention sufficiently to impair spatial memory, suggesting that overdependence on GPS devices hinders how we build spatial mental representations of our environments (Gardony, Brunyé, & Taylor, 2015). As described by Coluccia (2007), by pivoting our attention and studying map learning processes, we can develop a better understanding of how we learn and remember new environments and how we may maximize our memory performance.

### The Present Studies

In the present study, I aim to investigate the role of metacognitive monitoring and study repetition in map learning. Regarding recall performance, I hypothesize main effects of study repetition and JOL production on memory accuracy. I also hypothesize an interaction of study repetition and JOL production, such that participants are more likely to recall judged restudied items than unjudged restudied items.

Building on previous literature, I predict to see a positive relationship between study repetition and JOL accuracy as well as a positive relationship between study repetition and recall performance. Specifically, judged, repeat-studied stimuli will yield greater JOL accuracy than judged, once-studied items. Moreover, judged, repeat-studied stimuli will yield greater percent recall than judged, once-studied stimuli. Unjudged, repeat-studied stimuli will yield lower percent recall, and unjudged, once-studied stimuli will yield the lowest percent recall.

Further, my experiment includes two different study contexts (i.e., map and word-pairs), but testing is in word-pair format. As informed by the TAP hypothesis (Morris et al., 1977), I hypothesize that stimuli studied in a non-spatial context will yield a greater percentage of recalled items during test.

My study adds to previous research bridging map learning and metacognition, for through deeper understanding of how we monitor our encoding processes during map learning, we can inform more effective metacognitive processes that optimize our performance when retrieving our memory of learned environments.

## Methods

### Participants

In reviewing previous metacognitive monitoring studies that employed a similar design, we aimed to recruit 72 participants total and randomly assign 36 participants to each judgment group (Mitchum, Kelley, & Fox, 2016; Janes, Rivers, & Dunlosky, 2018; Myers, Rhodes, & Hausman, 2020). Ultimately, 37 Tufts University undergraduate introductory psychology students participated in our study, and we randomly assigned 20 participants to the JOL group and 17 participants to the RNG group. We excluded one participant response given that they demonstrated inattention during the task's attention checks.

Participants consisted of 59.5% females and 37.8% males, whereas 2.7% preferred not to state their gender. Their ages ranged from 18 to 22, with a mean age of 19. The participants' racial makeup was 8.1% Hispanic, 24.3% Asian, 10.8% Black or African American, 59.5% White, and 10.8% Other. Because some participants selected more than one race, their responses sum to greater than 100%.

## Design

Our experimental design employed a 2 x 2 x 2 mixed design, where we manipulated metacognitive monitoring (delayed JOL, delayed RNG), study context (map, word pair), and stimulus repetitions (one presentation, three presentations). We manipulated metacognitive monitoring (MM) as a between-participants variable by assigning participants to either produce a delayed judgment of learning (JOL) or complete a delayed random-number generation (RNG) control task following study of all landmark pair stimuli. In other words, half of the participants made JOLs and half of the participants completed RNG tasks. We included the RNG control group to ensure that we can assess the impact of prediction on memory (Rhodes, 2016).

We adjusted stimulus repetition on a within-participants basis, where all participants viewed half the landmark pairs once and half the landmark pairs three times. We implemented discrete spatial and non-spatial landmark pair stimuli contexts, such that all participants viewed one block of landmarks in a map context and another block in a listed word-pair context. Stimulus repetition and study context served as within-participant variables (see Table 1).

As a counterbalance measure, we controlled for context order so participants either viewed the map- or list-context pairs first. We also counterbalanced the assignment of restudied pairs such that when participants study pair stimuli 1-5 three times for Landmarks A, they study pair stimuli 6-10 three times for Landmarks B, and vice versa.

## Materials

### *Stimuli*

Stimuli consisted of two lists of 10 landmark pair stimuli– Landmarks A and Landmarks B (see Table 2). Landmark names were general (e.g., movie theatre) and did not possess specific brands (e.g., Ghirardelli Square). Participants studied a given landmark pair in the map context as two discrete locations that were marked by green stars and labeled with their landmark names (see Figure 1). In contrast, participants studied a given landmark pair in the list context as an associated word pair (see Figure 2).

To counterbalance stimuli spatial contexts, we randomly assigned half the participants to study Landmarks A in the map context and Landmarks B in the list context and, conversely, half the participants to study Landmarks A in the list context and Landmarks B in the map context. Participants studied all landmark pairs within one spatial context before proceeding to study the landmark pairs of the contrasting spatial context. Whereas half the participants studied map-context stimuli first, the other half studied list-context stimuli first (see Table 2).

To achieve two levels of study repetition, we designed our stimuli such that within each 10-item list, participants studied five landmark pairs 1 time each and five other landmark pairs 3 times each. Within these closed landmark lists, we spaced apart repeated landmark pairs such that on average, six trials intervene between each presentation of a particular three-study map landmark pair (*SD* = 0.833). Similarly, on average, six trials intervene between each presentation of a particular three-study list landmark pair *(SD* = 0.868).

## Procedure

### *Practice Phase*

At the start of the experiment, participants completed practice trials to familiarize themselves with the study's protocol. Participants viewed one map-context landmark pair and one list-context landmark pair, and both differed from those included in the study phase. For each studied landmark pair, participants studied the first landmark (cue landmark) alone for 200 milliseconds. Then, participants studied the target landmark for 200 milliseconds. Lastly, participants studied the cue and target landmarks together for 200 milliseconds. After studying the landmark pairs, participants engaged in their assigned MM group. For each landmark pair, all participants– that is, participants in both in the JOL and RNG condition– viewed the first landmark (cue landmark) alone.

Half of the participants observed the question, "How likely are you to remember the second landmark when presented the first if tested at a later time?" and inputted their JOL ratings in the following fashion: "1 not likely to remember/ 2 somewhat likely to remember/ 3 likely to remember/ 4 highly likely to remember." Meanwhile, the other half of the participants carried out the RNG task: "Pick a random number between 1 and 4," in which they inputted their selected random number. After studying and rating the two practice stimuli, participants completed the practice cued-recall test, where for each studied landmark pair, participants read the cue landmark and inputted the associated target landmark. If participants could not guess the name of a target landmark, they submitted an "X" in the answer box, as listed in the instructions. Participants provided their responses using a computer keyboard. Following the practice cued-recall test, participants viewed feedback on their recall performance. Specifically, they observed the two practice test questions' correct answers.

### *Study Phase*

During the study phase, participants studied the two landmark pair lists in succession. For each list, participants first viewed the full set of landmarks in their respective context for 30 seconds. To be specific, for map-context landmark pairs, they studied 10 unassociated landmark locations on a map. For verbal-context landmark pairs, they studied a list of 10 unpaired landmarks. Then, participants studied each landmark pair individually for 6 seconds. After viewing all landmark pairs at their assigned study repetitions, participants completed an 8-minute retention interval task before engaging in their assigned MM level.

**Landmark Pair Study.** Participants viewed their first 10 landmarks for 30 seconds. If participants received the map-context-first condition, they studied a map of 10 unassociated landmarks. Otherwise, they studied a list of 10 unpaired landmarks. When 30 seconds elapsed, participants advanced to closed landmark pair study. Stimulus presentation during the study phase followed the presentation procedures used during the practice phase. Given that participants studied five landmark pairs 1 time and five landmark pairs 3 times, participants completed 20 study trials for the first stimuli list.

When participants studied all 10 landmark pairs at their target study repetition levels, participants proceeded without break to study the remaining 10 landmark pairs in the second list. The second list began identically to the first block, initially with 30 seconds of viewing all 10 landmarks in their respective spatial context. Then, participants studied the last 10 landmark pairs one at a time in the same fashion as described above. At the end of studying their second stimuli list, participants have completed 40 study trials total.

**Retention Interval.** Immediately after studying the 20 landmark pairs to their targeted levels of study repetition, participants performed four 2-minute hidden pictures tasks, which involved visual search. For each hidden picture task, participants searched an illustrated image for hidden items portrayed in the margins. After viewing the illustrated image for two minutes, participants reported how many target items they found. Participants provided their responses using a computer keyboard. After participants completed all four hidden picture tasks, they proceeded to the next phase of the experiment.

**Metacognitive Monitoring.** During the rating phase, participants engaged in their assigned MM group. Worded differently, for each landmark pair, all participants viewed the cue landmark alone. While they viewed the cue landmark, half of the participants produced JOLs where they rated their likeliness from 1 to 4 (least to most likely) of successful recalling the associated target landmark during future test. They responded to the following question: “How likely are you to remember the second landmark when presented the first if tested at a later time?” The participants responded with one of the following options: “1 not likely to remember/ 2 somewhat likely to remember/ 3 likely to remember/ 4 highly likely to remember.”

Meanwhile, while they viewed the cue landmark, the other half of the participants generated a random between 1 and 4, as prompted by the instructions, “Pick a random number between 1 and 4.” Participants provided their responses using a computer keyboard.

### *Test Phase*

During the test phase, participants engaged in a closed-set cued-recall where they demonstrated memory for the 20 landmark pairs in the order of the lists they studied. For instance, participants who studied Landmarks A first completed the cued-recall test on Landmarks A first and Landmarks B second. One by one for each landmark pair, participants observed the name of the cue landmark and inputted the name of the associated target landmark. If participants could not guess the name of an associated target landmark, they provided an “X” in the answer box, as listed in the instructions. Participants provided their responses using a computer keyboard.

# Results

## Santa Barbara Sense of Direction (SBSOD) and Sense of Direction (SOD) Ratings

The JOL group yielded mean scores of 4.09 for the SBSOD Test and 3.45 for the self-reported SOD Rating. On the other hand, the RNG group yielded mean scores of 3.72 for the SBSOD Test and 3.18 for the self-reported SOD Rating.

## Analyzing Main Effects Using Linear Mixed Models

Using linear-mixed effects models, we examined the effects of MM group (JOL, RNG), study context (map, word-pair), and number of stimulus repetitions (one presentation, three presentations) on both memory accuracy and recall attempts. Specifically, we used the "lme4" package in R version 4.0.3 (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2020). The models for our analyses used a full factorial design and accounted for random effects of both participant and stimulus. This model took the following form:

DV ~ Rating Group (JOL, RNG) * Trial Context (map, word pairs) * Stimulus Presentations (1, 3) + (1|subject) + (1|stimulus) + ε

### *Memory Accuracy*

The results showed a significant main effect of study context (*odds ratio* = 0.68, *CI* = 0.56-0.83, $p < 0.001$) on memory accuracy. As visualized in Table 3, this outcome suggests that participants were 0.68 times less likely to accurately remember the target landmark if they studied the landmark pair in the map context (*M* = 0.237, *SD* = 0.426) compared to the word pair context (*M* = 0.357, *SD* = 0.48).

Additionally, the results demonstrated an interaction effect of MM group and study repetition on recall accuracy that approaches significance ($p = 0.08$), such that the RNG group's mean recall accuracy was greater for stimuli presented three times than stimuli presented once. Meanwhile, the JOL group's mean recall accuracy did not differ between stimuli presented once and stimuli presented three times (see Table 4). The JOL group exceeded the RNG group's mean recall accuracy for both levels of stimulus presentation. Interestingly, the RNG group's mean recall accuracy for stimuli presented three times (*M* = 0.313, *SD* = 0.12) is near that of the JOL group's for stimuli presented three times (*M* = 0.32, *SD* = 0.105).

### ***Recall Attempts***

We defined recall attempts as test responses where participants guessed the name of the associated target landmark, irrespective of the guesses' accuracy. In other words, we marked participants' responses as recall attempts when they inputted a landmark name. We did not mark participants' responses as recall attempts when they inputted an "X," an option given to participants in the instructions.

Similar to our results on recall accuracy, we saw a significant main effect of study context (*odds ratio* = 0.65, *CI* = 0.54-0.79, $p < 0.001$) on recall attempts (see Figure 5). This effect suggests that participants were 0.65 times less likely to attempt recalling the target landmark if they studied the pair in the map context (*M* = 0.571, *SD* = 0.496) compared to the word pair context (*M* = 0.706, *SD* = 0.456).

Moreover, the results showed a significant interaction effect of MM group and stimulus repetition on recall attempts ($p = 0.016$). In particular, the RNG group demonstrated greater mean recall attempts for stimuli presented three times than stimuli presented once. Conversely, the JOL group made fewer recall attempts for stimuli presented three times compared to stimuli presented once (see Figure 6). The RNG group's mean recall attempts for stimuli presented three times ($M = 0.673$, $SE = 0.122$) slightly exceeded the JOL group's mean recall attempts for stimuli presented once ($M = 0.67$, $SE = 0.105$).

Lastly, the results showed a significant three-way interaction effect of MM group, study context, and stimulus repetition ($p = 0.046$) on mean recall attempts (see Figure 7). Specifically, for stimuli studied in map context, the JOL group's mean recall attempts did not differ between stimuli presented once and stimuli presented three times. For stimuli studied in word pair context, the JOL group showed greater mean recall attempts for stimuli they studied once compared to stimuli they studied three times. Similarly, for stimuli studied in map context, the RNG group's mean recall attempts did not differ between stimuli presented once and stimuli presented three times. On the other hand, for stimuli studied in word pair context, the RNG group showed greater mean recall attempts for stimuli they studied three times compared to stimuli they studied once. Notably, the RNG group's mean recall attempts for word-pair stimuli presented three times ($M = 0.76$, $SE = 0.111$) approached the JOL group's mean recall attempts for word-pair stimuli presented once ($M = 0.77$, $SE = 0.095$).

## Measuring JOL Resolution Using Gamma Correlations

To examine the relationship between predicted and actual memory performance, we conducted gamma correlations (Masson & Rothello, 2009). In particular, we separately computed gamma correlations between memory performance and the ratings given for JOL and RNG groups. By conducting *t*-tests and comparing whether the gamma correlations differed from zero, we measured *JOL resolution*: the strength of the relative relationship between JOL ratings and memory performance. We observed that JOL correlations with accuracy marginally differed from zero ($t(17)=2.00$, $p = 0.061$), which was not true for RNG correlations ($t(14) = -0.201$, $p = 0.844$). As expected, JOL ratings predicted actual memory performance, whereas RNG ratings were unrelated to actual memory performance.

Lastly, we examined gamma correlations separately for stimulus repetition (one presentation, three presentations) and also separately for the study contexts (map, word pairs). We observed that JOL correlations with accuracy given three stimulus presentations yielded greater significance ($p < 0.0001$) than JOL correlations with accuracy given one stimulus presentation ($p < 0.001$). For gamma correlation analyses with respect to each MM group's study context and study repetition, refer to Table 5.

## Discussion

In the present project, I aimed to investigate the influence of metacognitive monitoring and study repetition in map learning. We used a JOL-reactivity, spaced rehearsal methodology to investigate the impact of JOL production and restudy on learning through maps and word pairs. In each context, participants studied half the landmark pairs once and half the landmark pairs three times. After studying all landmark pairs, half of the participants produced delayed JOLs to studied items while the other half completed delayed random number generation (RNG), also in context of the studied items. Participants lastly tested their memory in a non-spatial context.

Our early and exploratory investigations suggest that one's spaced rehearsal and monitoring of memory influences how they encode and retrieve landmark information. In the following section, I evaluate my initial hypotheses given our statistical analysis outcomes.

## Highlights of the Present Study

### *The Effect of Study Repetition on JOL Resolution and Memory Accuracy*

We found that participants produced more accurate JOLs for stimuli they studied three times compared to those they studied once. Consistent with our initial hypotheses, as well as previous studies, our results suggest that multiple study repetitions yield more reliable predictions of future retrievability (Logan et al., 2012; Pyc & Rawson, 2012). As expected, the RNG group recalled stimuli they studied three times more accurately than stimuli they studied once. In fact, the RNG group recalled stimuli they studied three times near the accuracy at which the JOL group recalled stimuli they studied once. The JOL group recalled stimuli with similar accuracy for both levels of stimulus presentation, with their mean recall accuracy for once-studied stimuli outperforming the RNG group's overall accuracy.

Although we did not detect a strong positive relationship between study repetition and recall performance, we found that study repetition and study context interacted to enhance memory, such that recall accuracy and JOL resolution was greatest when participants studied stimuli three times in the same spatial context that is present during test.

### *The Effect of Study Context on Memory Accuracy*

As expected, study context significantly affected memory accuracy. Participants in both JOL and RNG groups recalled stimuli they studied in the non-spatial, word pair context more accurately than stimuli they studied in the spatial map context. Such results align with Morris and researchers' Transfer Appropriate Processing (TAP) hypothesis (1977). That is, because participants used shared cognitive processes when learning and recalling non-spatial context stimuli, they recalled non-spatial context stimuli at a greater rate than spatial context stimuli.

### *Exploratory Analyses of Memory Attempts*

Though I did not make predictions about recall attempts, our exploratory analyses for offer early insight into how JOL production, study context, and study repetition influence participants' memory activation. Specifically, we observed that both MM groups attempted to recall more stimuli they studied in the non-spatial word pair context than those they studied in the spatial map context, thus supplementing our evidence for the TAP hypothesis (1977). This pattern suggests that when participants learned stimuli in a non-spatial context, they accessed similar cognitive resources during test and therefore facilitated different memory activation than when study and test contexts differ from each other.

We also noted that the RNG group made more recall attempts for stimuli they studied three times than for stimuli they studied one time. In contrast, the JOL group made fewer recall attempts for stimuli they studied three times than for stimuli they studied one time. To be specific, the RNG group made recall attempts for stimuli presented three times at a level greater than the JOL group's mean recall attempts for stimuli presented once.

However exploratory, our recall attempt analyses propose the idea that study repetition affects how the JOL group evaluates potential responses differently from the RNG group, in that JOL participants input guessed answers for repeat-studied stimuli more selectively (i.e., less frequently) and RNG participants input guessed answers for repeat-studied stimuli less selectively (i.e., more frequently).

## Limitations

The present study has discussed how metacognitive monitoring and study repetition serves to direct metamemory accuracy and memory performance; now, it is essential for us to consider how we can improve our experimental design in the future.

### ***Statistical Power***

In reviewing previous metacognitive monitoring studies that employed a similar design, we strived to recruit 72 participants total and randomly assign 36 participants to each judgment group (Mitchum, Kelley, & Fox, 2016; Janes, Rivers, & Dunlosky, 2018; Myers, Rhodes, & Hausman, 2020). However, due to data collection time constraints, we received half the participation we desired, thus splitting our statistical power. Given that our JOL and RNG groups were at 55% and 44%, respectively, of our desired participation levels, we can benefit from recruiting more subjects in the future to enhance the power of our data analyses.

### *Remote Data Collection*

Because participants completed our study asynchronously and remotely, they did not have the opportunity to ask experimenters questions about the task's instructions in real-time. Given that only one participant failed an assigned attention check, it is likely that participants generally paid attention to listed instructions and question prompts. However, with experimenter presence and availability to answer clarifying questions, we can work to better ensure that when we re-implement our experimental design in the future, all participants are attentive during every task stage.

## Future Directions

Now that we have interpreted our results and study limitations, I next promote future research directions that will benefit our understanding of how metacognitive monitoring and restudy affect map learning.

### *Analyzing Santa Barbara Sense of Direction (SBSOD) and Sense of Direction (SOD) Scores*

Previous research has observed significant diversity in human map learning skill levels. Pazzaglia and De Beni (2001) showed how the spatial format (map, verbal) of instructions differentially influences landmark-centered and survey-oriented groups. Similarly, Wen, Ishikawa, and Sato (2011) demonstrated how spatial learners contrast in their learning of survey information due to their differing senses of direction– as demonstrated by their self-reported SOD scores– and distinct recruitment of verbal, visual, and spatial working memory.

As such, it is important for future metamemory studies to consider the role of individual perspective preference survey learning ability in map learning.

In the present study, we collected and measured participants' SBSOD scores and self-reported SOD ratings, which correlate strongly with spatial learning skill sets (Hegarty et al., 2002). In additional analyses, our participants' self-report SBSOD and SOD scores will help us correlate the relationship between initial spatial ability, metacognition, and map learning. Moreover, by analyzing participants' self-reported, subjective reflections on their task performances, we can learn more about their individual experiences of how producing JOLs influenced their efforts to learn through maps.

### ***Characterizing Cognitive Load Incurred by MM of Map Learning***

In future experiments investigating the role of metamemory and restudy on map learning, researchers can gain more insight by considering how individual metacognitive self-regulation abilities affects the kind of cognitive load incurred by map learning JOLs. For instance, Scott and Schwartz (2007) utilized the Inventory of Metacognitive Self-Regulation (IMSR) (Howard, McGee, Shia, & Hong, 2000), which is a 32-item self-report questionnaire to calculate participants' metacognitive knowledge and abilities. By accounting for individual metacognitive capacities, researchers can more effectively characterize adaptive metacognitive monitoring methods to incur high germane cognitive load and improve spatial memory performance.

## Concluding Remarks

My study builds onto prior research bridging metacognition and map learning. By understanding how we can accurately monitor our encoding processes during map learning, we can inform more effective metamemory strategy and enhance our map learning memory performance. Spatial researchers have demonstrated how using a navigational aid splits one's attention sufficiently to impair spatial memory (Gardony, Brunyé, & Taylor, 2015). Accordingly, it is necessary for researchers to investigate different methods that can better shape self-motivated monitoring (i.e., through judging our learning) and control (i.e., through restudying select items) to optimize our learning through maps. By understanding the role of metamemory and restudy in map learning processes, we can better monitor our learning of new environments to shape more accurate map recollection during future test.

## References

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*(3), 327-355.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, *3*(3), 149-210.

Coluccia, E. (2007). Learning from maps: the role of visuo-spatial working memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*(2), 217-233.

Double, Kit S., Birney, Damian P., & Walker, Sarah A. (2018) A meta-analysis and systematic review of reactivity to judgements of learning, *Memory*, *26*(6), 741-750.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*(4), 374-380.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*(4), 551–565.

Gardony, A.L., Brunyé, T.T., & Taylor, H.A. (2015). Navigational aids and spatial memory impairment: The role of divided attention. *Spatial Cognition & Computation. 15(4),* 246-284.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151-176.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425-447.

Howard, B. C., McGee, S., Shia, R., & Hong, N. S. (2000). Metacognitive Self-Regulation and Problem-Solving: Expanding the Theory Base through Factor Analysis.

Kelemen, W. L., & Weaver, C. (1997). Enhanced memory at delays: Why do judgments of learning improve over time? Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 1394–1409.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. Journal of Memory and Language, 52, 478–492.

Kulhavy, R. W., Pridemore, D. R., and Stock, W. A. (1992). Cartographic experience and think aloud about thematic maps. *Cartographica* 291: 1–9.

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, *7*(3), 175-195.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*(2), 200.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Nilsson, R. M., & Mayer, R. E. (2002). The effects of graphics organizers giving cues to the structure of a hypertext document on users' navigation strategies and performance. International Journal of Human-Computer Studies, 57, 1-26.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. Educational Psychologist, 38(1), 1-4.

Pazzaglia, F., & De Beni, R. (2001). Strategies of processing spatial information in survey and landmark-centred individuals. *European journal of cognitive psychology*, *13*(4), 493-508.

Pazzaglia, F., & Taylor, H. A. (2007). Perspective, instruction, and cognitive style in spatial representation of a virtual environment. *Spatial Cognition and Computation*, *7*(4), 349-364.

Pyc, M. A., & Rawson, K. A. (2012). Are judgments of learning made after correct responses during retrieval practice sensitive to lag and criterion level effects?. *Memory & Cognition*, *40*(6), 976-988.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory.

Scott, B. M., & Schwartz, N. H. (2007). Navigational spatial displays: The role of metacognition as cognitive load. *Learning and Instruction*, *17*(1), 89-105.

Siegel, A. W., Krasic, K. C., & Kail, R. V. (1978). Stalking the elusive cognitive map. In *Children and the Environment* (pp. 223-258). Springer, Boston, MA.

Stamm, J., Johnson, J. T., and Verdi, M. P. (1999). The importance of structural map lines on recalling text information. Paper presented at *The Annual Meeting of the American Educational Research Association*, Montreal, Canada.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251-296.

Taylor, H. A., Naylor, S. J., & Chechile, N. A. (1999). Goal-specific influences on the representation of spatial perspective. *Memory & cognition*, *27*(2), 309-319.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive psychology*, *14*(4), 560-589.

Verdi, M. P., & Kulhavy, R. W. (2002). Learning with maps and texts: An overview. *Educational psychology review*, *14*(1), 27-46.

Wen, W., Ishikawa, T., & Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. *Applied Cognitive Psychology*, *25*(4), 654-662.

## Appendix

**Table 1**

*Table Showing Experiment's Manipulations of MM and Stimulus Repetition*

| | | Spatial landmark pairs presented three times | |
|---|---|---|---|
| | | Landmark pairs 1-5 | Landmark pairs 6-10 |
| Judgment | JOL | Landmarks A in map context, presented first | Landmarks B in map context, presented first |
| | | Landmarks B in map context, presented second | Landmarks A in map context, presented second |
| | RNG | Landmarks B in map context, presented first | Landmarks A in map context, presented first |
| | | Landmarks A in map context, presented second | Landmarks B in map context, presented second |

*Note*. When participants study pair stimuli 1-5 three times for Landmarks A, they study pair stimuli 6-10 for Landmarks B, and vice versa.

**Table 2**

*Landmark Pair Stimuli Lists Used in Experiment*

| | Landmarks A | | | Landmarks B | |
|---|---|---|---|---|---|
| 1 | Movie Theatre | Zoo | 1 | Arcade | Bakery |
| 2 | Bank | Auditorium | 2 | Grocery Store | Tennis Court |
| 3 | Pharmacy | Art Gallery | 3 | ATM | Synagogue |
| 4 | Car Wash | Post Office | 4 | Auto Shop | Pool |
| 5 | Train Station | Realtor | 5 | Bus Stop | Garden |
| 6 | Gym | Hotel | 6 | Library | Pub |
| 7 | Nail Salon | School | 7 | Barber | Hospital |
| 8 | Restaurant | Laundromat | 8 | Courthouse | Factory |
| 9 | Golf Course | Apartment | 9 | Travel Agent | Museum |
| 10 | Church | Flower Shop | 10 | Motor Home | Aquarium |

**Table 3**

*Effects of MM Level, Study Context, and Study Repetition on Memory Accuracy*

| | **Recall Accuracy** | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 0.31 | 0.19 – 0.51 | **<0.001** |
| MM Level | 1.22 | 0.75 – 1.98 | 0.426 |
| Study Context | 0.68 | 0.56 – 0.83 | **<0.001** |
| Study Repetition [1] | 0.86 | 0.71 – 1.04 | 0.120 |
| MM Level * Study Context | 0.99 | 0.82 – 1.21 | 0.940 |
| MM Level * Study Repetition [1] | 1.19 | 0.98 – 1.45 | 0.080 |
| Study Context * Study Repetition [1] | 1.11 | 0.92 – 1.35 | 0.273 |
| (MM Level * Study Context) * Study Repetition [1] | 0.94 | 0.78 – 1.15 | 0.554 |

**Table 4**

*Effects of MM Level, Study Context, and Study Repetition on Recall Attempts*

| | **Recall attempts** | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 2.54 | 1.39 – 4.66 | **0.003** |
| MM Level | 1.19 | 0.65 – 2.17 | 0.575 |
| Study Context | 0.65 | 0.54 – 0.79 | **<0.001** |
| Study Repetition [1] | 0.91 | 0.76 – 1.10 | 0.348 |
| MM Level * Study Context | 0.90 | 0.74 – 1.09 | 0.287 |
| MM Level * Study Repetition [1] | 1.26 | 1.04 – 1.53 | **0.016** |
| Study Context * Study Repetition [1] | 1.05 | 0.87 – 1.27 | 0.598 |
| (MM Level * Study Context) * Study Repetition [1] | 0.82 | 0.68 – 1.00 | **0.046** |

**Table 5**

*Gamma Correlations (*p*-value) as a Function of MM Level, Study Context, and Study Repetition*

| MM | Overall | One study repetition | Three study repetitions | Map Context | Word Pair Context |
|---|---|---|---|---|---|
| JOL | 0.394<br>(0.06) | 0.848<br>(**<0.001**) | 0.849<br>(**<0.0001**) | 0.813<br>(**<0.01**) | 0.454<br>(**0.04**) |
| RNG | 0.023<br>(0.844) | 0.013<br>(0.933) | 0.052<br>(0.752) | 0.020<br>(0.878) | 0.103<br>(0.417) |

**Figure 1**. *Example of Map Context Landmark Pair Used in Experiment*. During study of the map context list, participants studied no more than two landmarks (i.e., one landmark pair) at a time. That is, participants never studied two landmark pairs concurrently.

Arcade --- Bakery

**Figure 2.** *Example of List Context Landmark Pair Used in Experiment*. During study of the list-context list, participants studied no more than two landmarks (i.e., one landmark pair) at a time. That is, participants never studied two landmark pairs concurrently.

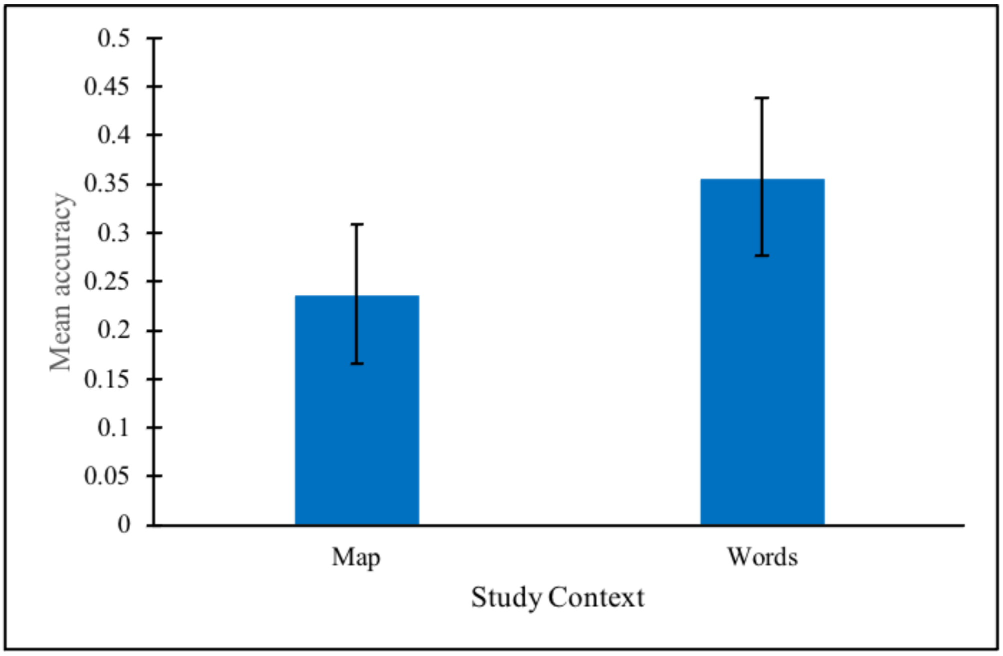

**Figure 3:** *The Effect of Study Context on Mean Recall Accuracy.* The error bars represent the standard error of the mean.

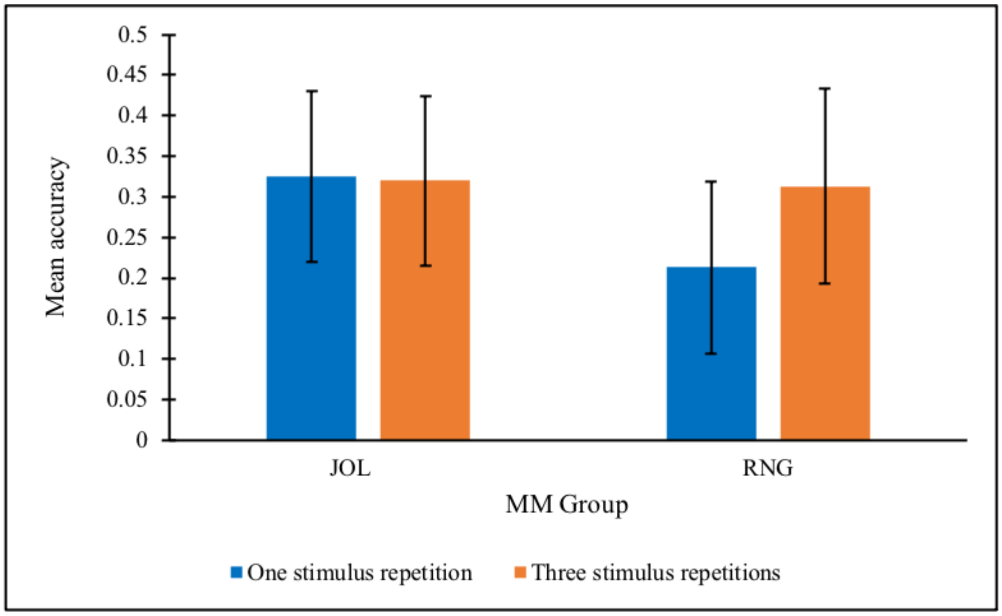

**Figure 4:** *Marginal Interaction Effect of Stimulus Repetition and MM Group on Mean Recall Accuracy.* The error bars represent the standard error of the mean.

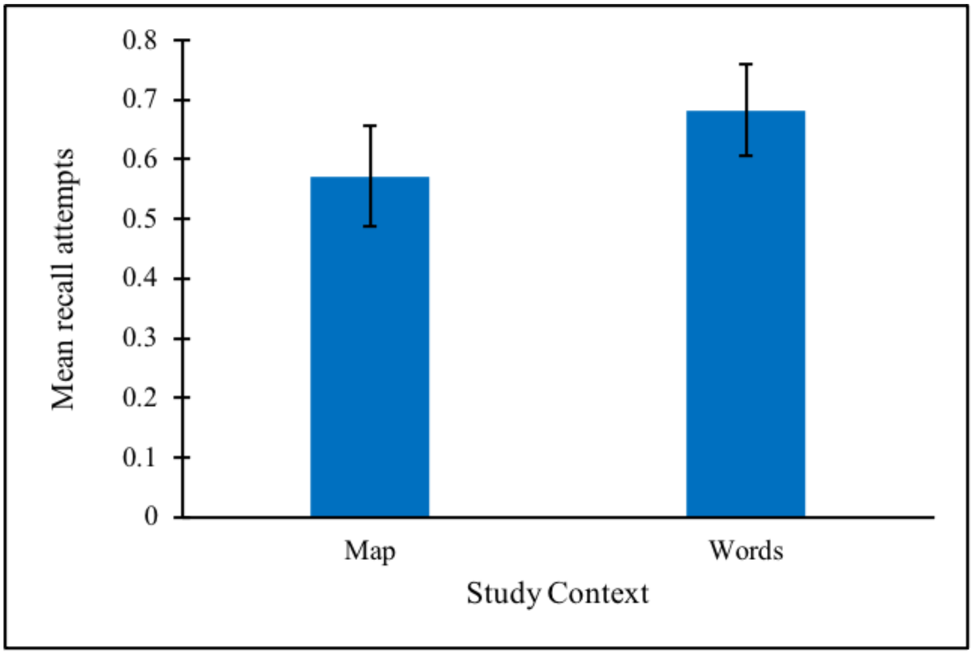

**Figure 5:** *The Effect of Study Context on Mean Recall Attempts.* The error bars represent the standard error of the mean.

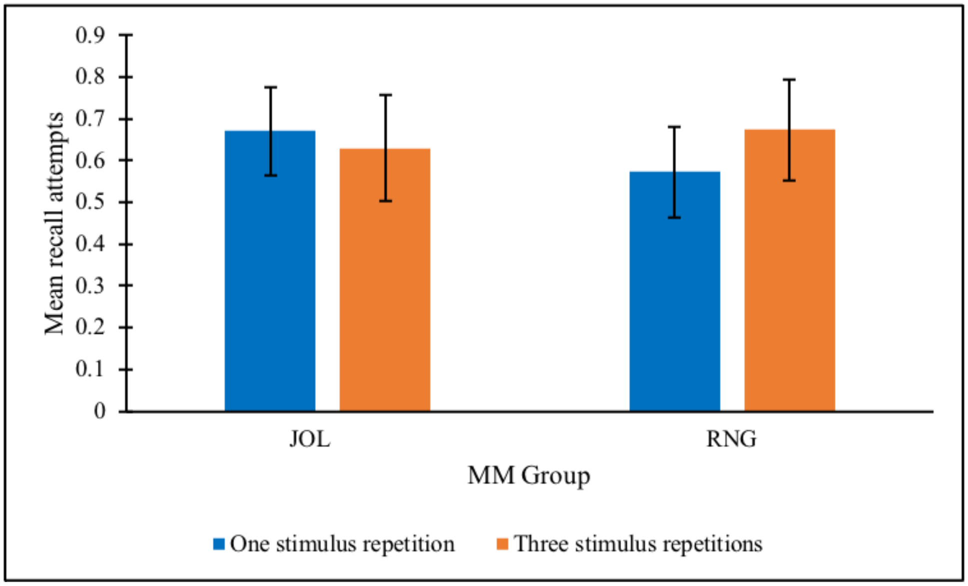

**Figure 6:** *Interaction Effect of Stimulus Repetition and MM Level on Mean Recall Attempts.* The error bars represent the standard error the mean.

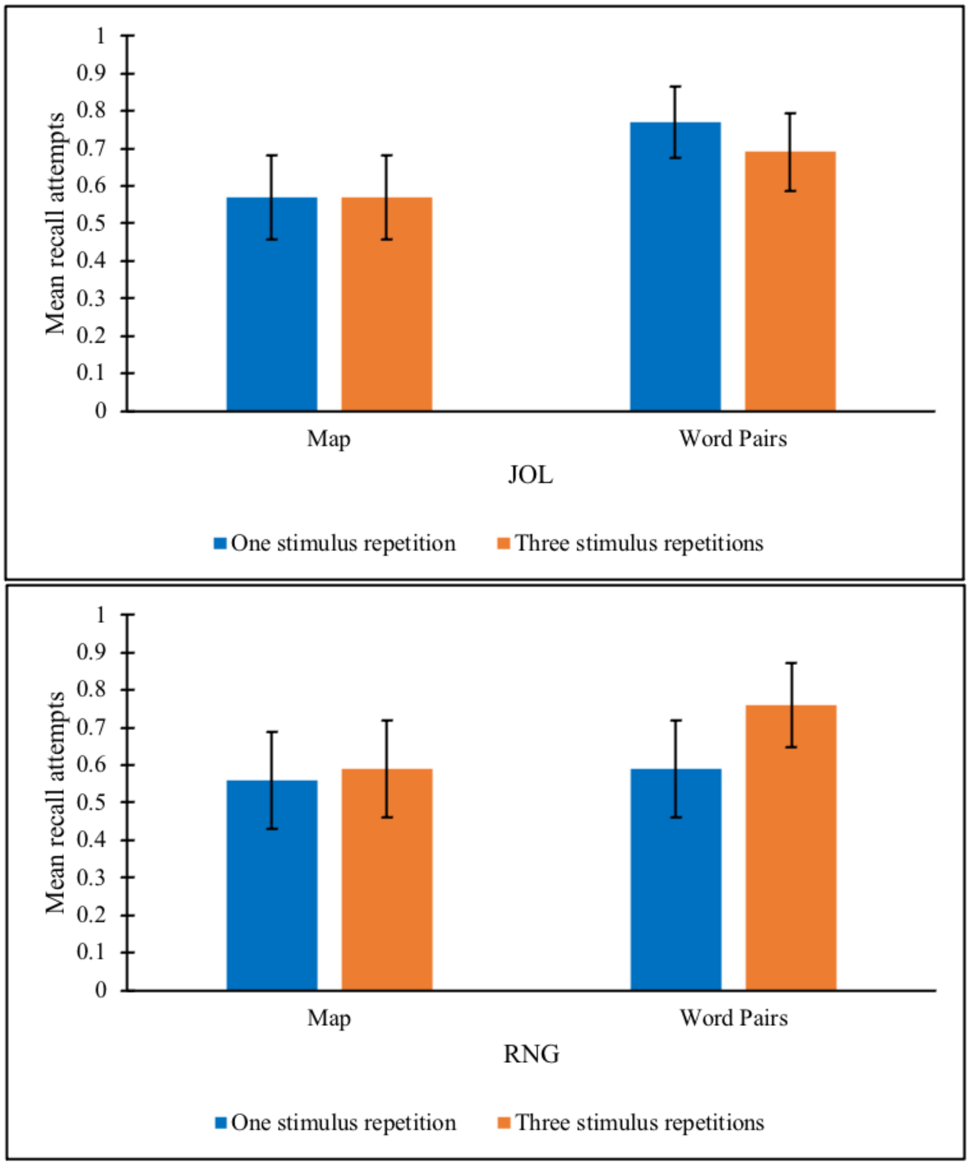

**Figure 7:** *Three-Way Interaction Effect of Study Context, Stimulus Repetition, and JOL Production on Mean Recall Attempts.* The error bars represent the standard error the mean.