THE PRECISION OF RETRIEVING TEMPORAL INFORMATION: BEHAVIORAL AND ELECTROPHYSIOLOGICAL STUDIES

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

JOHN E. SCOFIELD, M.S.

Dr. Jeffrey D. Johnson, Dissertation Supervisor

DECEMBER 2020

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE PRECISION OF RETRIEVING TEMPORAL INFORMATION: BEHAVIORAL AND ELECTROPHYSIOLOGICAL STUDIES

presented by John E. Scofield, M.S.,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Associate Professor Jeffrey Johnson (Chair)

Curators' Distinguished Professor David Geary

Curators' Distinguished Professor Nelson Cowan

Assistant Professor Roxana Botezatu

ACKNOWLEDGMENTS

I would like to first thank my advisors Dr. Jeffrey Johnson and Dr. David Geary for their continual support and guidance throughout my graduate education. The training and mentorship from Drs. Johnson and Geary serve as an invaluable resource as I continue to utilize skills as a research scientist moving forward and throughout future careers. I would second like to thank the remainder of my committee members, Drs. Cowan and Botezatu for their quality time, feedback, and support throughout this process. Lastly, I extend gratitude to the Department of Psychological Sciences and the faculty and staff, without which my education would not have been possible.

TABLE OF CONTENTS

ACKNOWLEDGMENTS……………………………………………………………..ii

LIST OF TABLES……………………………………………………………….v

LIST OF FIGURES…..…………………………………………………….. …v-vi

ABSTRACT…………………………………………………………………….vii

Chapter

1. GENERAL INTRODUCTION…………………………………………………...1
    1.1 The Current Experiments…………………………………………….5
2. EXPERIMENT 1……………………………………………………………….8
    2.1 METHOD…………………………………………………………...10
        2.1.1 Subjects……………………………………………………….10
        2.1.2 Stimuli…………………………………………………………11
        2.1.3 Procedure……………………………………………………..11
        2.1.4 Data Analysis…………………………………………………13
    2.2 RESULTS AND DISCUSSION……………………………………..14
        2.2.1 Changes in Temporal Precision Across Time………………..16
        2.2.2 Response Bias in Temporal Judgments……………………...17
        2.2.3 Guessing in Temporal Judgments……………………………19
        2.2.4 Confidence Judgments………………………………………20
3. EXPERIMENT 2……………………………………………………………..22
    3.1 METHOD…………………………………………………………...23
        3.1.1 Subjects……………………………………………………….23
        3.1.2 Stimuli and Procedure……………………………………….23
        3.1.3 Data Analysis…………………………………………………24
    3.2 RESULTS AND DISCUSSION……………………………………..24
        3.2.1 Standard Mixture-Modeling…………………………………27
        3.2.2 Trial-Level Modeling of Recency Effects……………………..30
4. EXPERIMENT 3……………………………………………………………..34
    4.1 METHOD…………………………………………………………...38

4.1.1 Subjects……………………………………………………………….38
4.1.2 Stimuli and Procedure……………………………………………….39
4.1.3 EEG Acquisition and Processing………………………………...40
4.1.4 Data Analysis…………………………………………………………41
4.2 RESULTS AND DISCUSSION………………………………………………..42
4.2.1 Behavioral Results……………………………………………………42
4.2.2 Event-Related Potentials……………………………………………48
4.2.3 Oscillatory Activity……………………………………………………52
5. EXPERIMENT 4……………………………………………………………………..57
5.1 METHOD…………………………………………………………………………...58
5.1.1 Subjects………………………………………………………………….58
5.1.2 Stimuli and Procedure……………………………………………….59
5.1.3 Data Analysis…………………………………………………………61
5.2 RESULTS AND DISCUSSION………………………………………………..62
5.2.1 Response Times………………………………………………………...62
5.2.2 Relating Spatial and Temporal Precision……………………….63
5.2.3 Recency Effects in Temporal and Spatial Precision………..69
6. GENERAL DISCUSSION…………………………………………………………..71

REFERENCES ……………………………………………………………………………..83
APPENDIX ………………………………………………………………………………...100
VITA ………………………………………………………………………………………..102

LIST OF TABLES

Table Page

1. Summary of Trial-Level Modeling Estimates for Experiment 2......................32

LIST OF FIGURES

Figure Page

1. Task Schematic for Experiment 1……………………………………………..13
2. Behavioral Results for Temporal Judgments in Experiment 1………………..16
3. Effects of Response Bias on Temporal Judgments in Experiment 1………...19
4. Summary of Temporal and Confidence Judgments in Experiment 2………..26
5. Behavioral Results for Temporal Judgments in Experiment 2………………..27
6. Mixture-Modeling Results for Experiment 2…………………………………….30
7. Trial-Level Modeling of Precision According to Animacy in Experiment 2…..33
8. Behavioral Results for Temporal Judgments in Experiment 3………………..44
9. Effects of Recency on Temporal Precision in Experiment 3………………….45
10. Effects of Response Bias and Guessing in Experiment 3……………………48
11. Left Parietal ERPs According to Precision Level in Experiment 3……………50
12. Left Parietal ERP Amplitudes by Precision Level in Experiment 3…………...51
13. Regression Results of Precision and ERP Amplitude in Experiment 3……...52
14. Regression Results of Precision and Frequency Power in the Alpha Band in Experiment 3……………………………………………………………………….54
15. Clustered Regression of Precision and Frequency Power in Experiment 3...55

16. Regression Results of Precision and Frequency Power in the Theta Band in Experiment 3……………………………………………………………………….56
17. Behavioral Response Times in Experiment 4…………………………………..63
18. Error Distributions of Temporal and Spatial Judgments in Experiment 4……64
19. Regression Results of Temporal and Spatial Precision in Experiment 4……65
20. Transformed Temporal and Spatial Precision Results in Experiment 4……..66
21. Regression Results of Temporal and Spatial Precision by Judgment Order in Experiment 4……………………………………………………………………….68
22. Subject-Wise Regression Results of Temporal and Spatial Precision by Judgment Order in Experiment 4……………………………………………..…..68
23. Recency Effects of Temporal and Spatial Precision in Experiment 4………..70

THE PRECISION OF RETRIEVING TEMPORAL INFORMATION: BEHAVIORAL AND ELECTROPHYSIOLOGICAL STUDIES

John E. Scofield, M.S.

Dr. Jeffrey D. Johnson, Dissertation Supervisor

ASTRACT

The knowledge of when an event took place provides benefits to episodic memory, such as distinguishing among multiple traces and learning event sequences. As a tool for understanding memory, time is appealing given its ever-changing quality, and the ease with which it is targeted at retrieval. Whereas studies of episodic retrieval typically employ categorical measures of retrieval, characterizing a continuous feature such as time warrants measures sensitive to the precision of retrieved information. Through four experiments, we adapted a paradigm for assessing the fine-grained precision of retrieval to understand the nature of judging the time at which a memory was encoded. Subjects studied a series of pictures and were subsequently tested on when they previously studied items. Temporal judgments were less accurate with passing time, with negligible guessing. Neurally, ERP amplitudes in left parietal electrodes tracked the precision of temporal judgments, with higher ERP amplitudes associated with better precision. Additionally, frequency power in both the alpha and theta bands were associated with temporal precision. Finally, while testing spatial retrieval, a correspondence emerged between spatial and temporal precision on a trial to trial basis, but a dissociation was found in which the recency effects found in temporal judgments was not present in spatial judgments. Together, these findings elucidate the role of time and space in episodic memory retrieval.

1. GENERAL INTRODUCTION

Episodic memory refers to the cognitive and neural processes that allow for the encoding and retrieval of memories for events and experiences that constitute our personal histories. A well-functioning episodic memory ability is additionally integral to a variety of cognitive abilities, enabling us, for instance, to flexibly navigate through familiar environments (Doeller, Barry, & Burgess, 2010), preserve a sense of autonoetic awareness of being able to remember the past (Tulving, 1983, 1985), and make predictions about the likelihood of future events taking place (Schacter, Addis, & Buckner, 2007; Tulving, 2002). One central feature of episodic memories is that they are thought to represent each individual event in a unique manner. This “uniqueness” can be supported by several factors, including the rich and distinct details surrounding the event, subjective aspects of retrieval such as vividness and confidence, and the ever-changing contextual information that situates events along both spatial and temporal dimensions. Whereas considerable effort has been devoted to characterizing many of these factors (for reviews, see Howard, 2017; Spaniol et al., 2009), the experiments comprising this dissertation were aimed at further understanding the role and characteristics of temporal information in episodic encoding and retrieval.

Experimental studies of episodic memory have predominantly relied on paradigms that require subjects to make categorical judgments about the basis of retrieval. For instance, the commonly employed *remember/know* task involves distinguishing with a binary judgment items on a memory test that are accompanied by the qualitative recollection of specific details (“remembering”) from those items that merely elicit an acontextual sense of familiarity (“knowing”; e.g., Mandler, 1980; Rajaram, 1993; Tulving, 1985; for reviews, see Skinner & Fernandes, 2007; Yonelinas, 2002). Such categorical responding is also the basis for the *source memory* paradigm (Johnson & Raye, 1981), in which the presence versus absence of retrieving a specific, experimentally-manipulated detail associated with test items is designated with forced-choice judgments (for reviews, see Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2009). Recent iterations of these tasks have moved toward incorporating measures that are more graded in nature, as is the case with having subjects indicate item familiarity by providing confidence ratings on a Likert-type scale (e.g., Montaldi, Spencer, Roberts, & Mayes, 2006; Yonelinas, Otten, Shaw, & Rugg, 2005). However, experimental investigations of the nature of recollection, and in particular whether it also has continuous properties, are much less common. A few notable exceptions to this come from studies that have probed subjects to indicate the amount of information they are able to retrieve for a given test item (e.g., Vilberg & Rugg, 2007; Wilding, 2000) or their graded level of confidence for making a source memory judgment (e.g., Hayes, Buchler, Stokes, Kragel, & Cabeza, 2011; Mickes, Wais, & Wixted, 2009; Woroch & Gonsalves, 2010).

The continuous variation with which we can retrieve and differentiate episodic memories has the potential to be of great utility, and is thus of growing interest, as it has been recently in the field of working memory research (e.g., Wilken & Ma, 2004; Zhang & Luck, 2008; for reviews of working memory and related issues, see Adams, Nguyen, & Cowan, 2018; Ma, Husain, & Bays, 2014). Bays, Catalao, and Husain (2009) have shown that when subjects use a continuous report task to judge the precise color of a previously studied item, memory precision exhibits graded effects with increases in working memory load. Moreover, within individuals, the precision of visual working memory representations can change from trial to trial, even in the absence of any particular experimental manipulation (Fougnie, Sucho, & Alvarez, 2012). These and other studies from the working memory have inspired other domains of memory research that the resolution, or *precision*, with which information can be retrieved has been demonstrated to vary continuously.

The continuous report task has recently been adopted in long-term memory studies to further understand the precision (or fidelity) with which episodic recollection occurs (e.g., Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Murray, Howie, & Donaldson, 2015). Many of these studies use tasks where spatial information serves as the continuous dimension of interest. In one initial study by Harlow and Donaldson (2013), subjects first encoded a series of words that were each presented along with a mark denoting a spatial position on a circle. Upon presentation of the word cue during a subsequent memory test, subjects were to mark the corresponding spatial position as precisely as possible. The main result was that the precision of the judgment varied continuously around the correct position (for similar findings, see Harlow & Yonelinas, 2016; Murray et al., 2015). Other researchers have conceptually replicated those findings using tasks in which spatial judgments across a two-dimensional area are estimated (Nilakatan, Bridge, Gagnon, VanHaerents, & Voss, 2017; Nilakatan, Bridge, VanHaerents, & Voss, 2018) and where color-wheel judgments as described above for working memory are employed (Brady et al., 2013). While the emerging pattern of findings from these studies points to declining precision while transitioning from shorter-term (working) to longer-term tasks, they importantly have moved episodic memory research from away from a solely thresholded treatment of recollection (see Parks & Yonelinas, 2009; Wixted, 2007).

Besides having the ability to detect more sensitively the finer changes in the quality of retrieved information, the continuous report task has also been beneficial in advancing a statistical modeling approach that can be used to characterize multiple decision processes involved in memory retrieval. This approach involves mixture modeling, which separates a distribution of memory responses into multiple components. The main component describes a distribution that represents the precision of memory retrieval that varies continuously around zero. This distribution is often a von Mises distribution with tasks requiring subjects to make judgments along a circular display, and can be other distributions depending on what scale is used (e.g., normal distributions can be used with linear scales). The width, or shape of this initial distribution generally characterizes the fidelity, or precision, in which subjects were able to retrieve items from memory. The other critical component of mixture-modeling is a secondary distribution that is superimposed on the distribution of test responses. This is a guessing parameter and is characterized by estimating the number of trials that best fit within a uniform distribution, such that there is equal probability that judgments are close or further away from the target. The guessing parameter represents the rate, or percentage of trials that subjects were estimated to have guessed on the trial, suggesting that retrieval had failed for a particular test item. Mixture-modeling following this framework is beneficial in that it not only yields continuous estimations of memory precision but can identify trials in which retrieval precision was irrelevant in the case of random guessing.

1.1 The Current Experiments

The set of four experiments reported on here were conducted to further characterize the processes supporting the retrieval of temporal information in episodic memory. Experiment 1 first provides an introduction on the importance and benefits of understanding temporal information retrieval. The experimental design is then described, constituting a novel adaptation of the continuous report task to responses made along a linear timeline. Subjects incidentally encoded an uninterrupted series of pictures and subsequently judged each item according to when, along the series, it was presented. This task, along with the associated mixture-modeling approach described above, allows for estimation of the precision of temporal retrieval judgments and the involvement of other decision processes (e.g., guessing and bias) in making such judgments. Furthermore, this analysis approach is used to test novel predictions about how temporal precision changes across the encoding period. A manuscript based on the results of Experiment 1 is available at https://psyarxiv.com/kfhqx/.

Experiments 2 and 3 next employed experimental designs that were largely similar to that of Experiment 1, but with a respective decrease and increase in the number of encoded and tested items. The primary aim of Experiment 2 was to provide a conceptual replication of the novel effects of recency on temporal precision observed in Experiment 1, with an *a priori* power analysis allowing for a reduction in the numbers of trials and subjects. Additionally, the stimuli employed in Experiment 2 were balanced across animate and inanimate categories (a design feature that was not considered at the time Experiment 1 was conducted), as previous research has indicated that memory performance can differ according to this distinction, which thereby could have affected our earlier estimation of temporal retrieval.

As outlined in the introduction for Experiment 3, the length of the encoding phase used during each of the preceding experiments raises the potential issue that responses are truncated by the start and end boundaries of the encoding list. This truncation could both limit the accuracy of estimating temporal retrieval precision and affect the degree to which effects of recency and bias are observed. Experiment 3 thus expanded on the first two experiments by considerably increasing the number of trials presented during the encoding phase. Due to the corresponding increase in the amount of time subjects were in the laboratory, a typical sample size was not practical. Instead, a small-*N* design was employed, and the opportunity was taken to acquire electroencephalographic (EEG) data from these subjects. Analysis of the EEG data, both in terms of event-related potentials (ERPs) and oscillatory activity, allowed for the novel investigation of the neural correlates of temporal memory precision.

Finally, Experiment 4 constituted a switch from the aforementioned laboratory protocols to an online behavioral setting, due to COVID-19. This study investigated the retrieval of temporal information in much the same way as previously described but also included an analogous task that probed the more commonly researched phenomenon of spatial memory retrieval. By having concurrent judgments for each of these tasks on every trial, dependencies between temporal and spatial retrieval, as well as similarities in precision and decision-making, could be directly tested.

2. EXPERIMENT 1

As described earlier, studies using continuous report tasks to investigate long-term retrieval precision have for the most part relied either on spatial judgments, where subjects might be required to encode and retrieve studied items associated with a marker positioned on a circle (e.g., Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Murray et al., 2015; Richter, Cooper, Bays, & Simons, 2016), or on judgments placed along a continuous color wheel (e.g., Brady et al., 2013). The degree to which temporal information is precisely retrieved, however, is much less understood. Whereas several classic memory studies involved asking subjects to estimate the time at which items were previously studied, such judgments were often relatively coarse in that they only distinguished between multiple encoding lists or between rather large sections of a single list (e.g., Hintzman & Block, 1971; McCormack, 1984). Some recent studies advanced this work by assessing temporal estimation at a higher resolution — that is, at the level of seconds or trials — for stimuli that were presented across a continuous encoding list (Jenkins & Ranganath, 2010; Lositsky et al., 2016; Montchal, Reagh, & Yassa, 2019). In one study, Montchal, Reagh, and Yassa (2019) used a rather unique experimental design in which subjects watched a half-hour popular television show and were later asked to place still-frames extracted from the show along a timeline indicating when it had occurred. Montchal et al. (2019) observed that subjects were highly accurate in their temporal judgments and that performance correlated, based on functional magnetic resonance imaging (fMRI) data, with activity in regions of the entorhinal cortex previously linked to processing temporal contextual information (Lositsky et al., 2016; Tsao et al., 2018). One notable aspect of these findings was that even subjects who did not watch the television show also exhibited some degree of accuracy at scene placement, suggesting that natural progression of the show's plot may have provided informative cues about time.

We built on the aforementioned findings in the current experiment by using an encoding phase that consisted of a series of pictures chosen randomly and presented at consistent intervals, thus minimizing the presence of any informative cues. This approach is somewhat comparable to that used in the study by Jenkins and Ranganath (2010), in which subjects were shown a series of pictures presented in the context of a working memory task and then later placed some of the items on a timeline corresponding to the encoding phase. The authors found that the items presented more recently (i.e. occurring later in the encoding phase) were associated with lower error rates in temporal estimation compared to the more remote items. However, based on the behavioral results, it is unclear whether other decision processes were involved in making the temporal judgments, as could be elucidated with mixture modeling. Additionally, because this study and Montchal et al. (2019) both involved the analysis of fMRI data, the neural correlates of temporal retrieval were identified in a relatively coarse manner by comparing groups of high- versus low-accuracy trials. It thus remains to be determined whether and how continuous changes in the precision of temporal retrieval would be borne out in neural differences.

Using a mixture-modeling approach with continuous behavioral judgments along a timeline corresponding to encoding, Experiment 1 was focused first on characterizing any changes in the precision of temporal retrieval across remote versus recent memories. Second, the modeling procedure allowed us to test for the influence of guessing on temporal memory. As noted earlier, guessing in the context of long-term memory studies employing continuous report tasks has been well described as uniform over a circular response scale (e.g., Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Murray et al., 2015; Richter et al., 2016). Because judgments about time require switching to a linear scale, particularly where one end of the scale does not lead continuously into the other end, it is an open question as to whether guessing plays a similar role in responding. Finally, we tested for the involvement of bias in temporal judgments which would take the form of a shift in the mean of the error distribution, reflecting a tendency to respond more recently or remotely relative to an item's actual encoding time. Given that judgments on the typically-employed circular scales have been shown to be largely unaffected by bias, the use of a linear scale in the present experiment might reveal a novel characteristic of decision-related processing when time is the response dimension.

2.1 METHOD

2.1.1 Subjects

Thirty-two University of Missouri (MU) students participated for partial course credit. Inclusion criteria were: 18-30 years old, native-English speaking, and no history of neurological disease. Informed consent was obtained in accordance with the MU Institutional Review Board. Data were initially screened to confirm each subject's active engagement with the task, defined as moving the computer mouse from the starting position on >95% of test trials (see Harlow & Yonelinas, 2016). Two subjects were excluded by this criterion, with an additional subject removed for having a high percentage (28%) of fast (< 500 ms) temporal judgments. The final sample of subjects (20 females, 9 males) were 18-22 years old (M = 19.1, SD = 1.1).

2.1.2 Stimuli

The stimulus pool consisted of 300 color pictures of common, nameable objects. Each picture subtended a visual angle of about 3.1° and was presented centrally on a light gray (~70% white) background. Stimuli were displayed on a 24-inch widescreen LCD monitor (1024×768 resolution) viewed at a distance of approximately 1 meter. Stimulus presentation was controlled with the Cogent2000 toolbox (v.1.32; http://www.vislab.ucl.ac.uk) in MATLAB (R2012a; MathWorks, Natick, MA).

2.1.3 Procedure

The experimental session (~45 minutes) consisted of an encoding phase followed by a test phase. Instructions and practice on encoding were administered first, keeping subjects naive about the nature of the test phase until immediately prior to its start. The encoding phase comprised a single block of 300 pictures, with each displayed for 3000 ms and followed by a central plus sign presented for 500 ms. Subjects were informed of the number of trials and total time of the encoding phase (17.5 minutes) prior to its start. Subjects were instructed to rate the pleasantness of each picture on a 4-point scale (very pleasant, somewhat pleasant, somewhat unpleasant, and very unpleasant) via key presses mapped to their left little through index fingers. The response options were displayed at the bottom of the screen throughout the encoding phase.

Following encoding, subjects received instructions and practice on the test phase, which consisted of all study pictures being presented in a randomized order. Each test trial began with a picture presented above a horizontal time scale. The time scale ranged from trial 1 on the left to trial 300 on the right, with labels marking every 50 trials. On each trial, an arrow was initially displayed at the midpoint of the scale (trial 150). Subjects were instructed to use the computer mouse to move the arrow, as precisely as possible, to the position at which they thought the picture occurred at study, and to click the left mouse button to finalize their response. A vertical confidence scale, ranging from 0% at the bottom to 100% at the top (with labels marking 20% increments), then replaced both the picture and time scale. An arrow next to the scale was initialized at 50% on each trial, and subjects were to rate their confidence about correctly retrieving the picture's study position, irrespective of precision (also see Harlow & Yonelinas, 2016). After the confidence judgment was made by pressing the left mouse button, a central plus sign was displayed for 1000 ms until the next trial began. Figure 1 shows a visualization of the experiment.

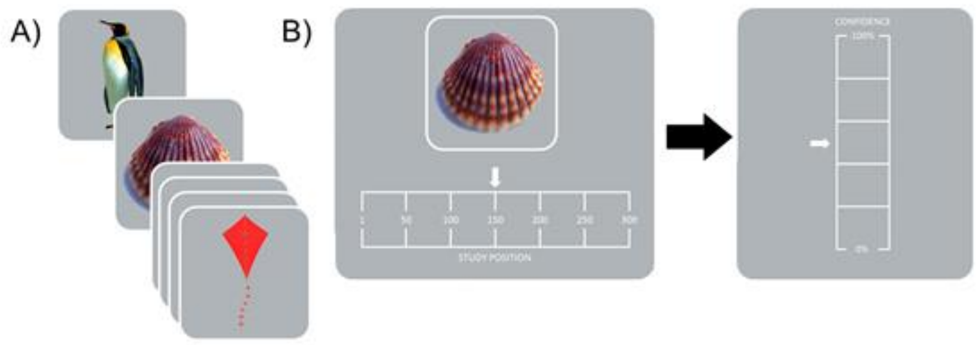

Figure 1. Task Schematic for Experiment 1
(A) Subjects first incidentally encoded 300 common objects that were each shown for 3000 ms. (B) During the test phase, subjects were shown all pictures again, along with a continuous scale representing the encoding phase timeline. Subjects moved the computer mouse horizontally to indicate, as precisely as possible, when the picture occurred. Subjects then rated their confidence about the preceding temporal judgment by moving the computer mouse vertically between 0 and 100 percent.

2.1.4 Data Analysis

The behavioral data and analysis scripts are available at https://osf.io/c7nge. Primary analyses were conducted using MemToolbox (v. 1.0.0; Suchow, Brady, Fougnie, & Alvarez, 2013) in MATLAB (R2012a), with modifications made to the toolbox to model responses along a linear (time) scale. Prior to analysis, study trials on which subjects failed to respond or did so outside a predefined interval of 500-3000 ms after picture onset were removed (2.9% of trials; SD = 3.0). Although the temporal position and confidence judgments during the test phase were self-paced, trials on which subjects made the temporal response faster than 500 ms or made either response slower than 10 s were removed (2.4% of trials; SD = 3.6). Given the infrequency of responses near the start and end of the scale, and due to the potential for temporal judgments to be affected by these boundaries, data for the first 50 and last 50 study trials were excluded from all analyses (for analogous approaches, see Montchal et al., 2019; Nilakantan et al., 2017, 2018).

Trials were scored in terms of error distance (in number of trials) between the actual study position of the picture and its judged position on the test. The temporal (position) scale was 600 pixels wide, providing a resolution of every .5 trial. The confidence scale was 500 pixels tall, giving a resolution of .2%. The posterior distribution of each parameter was estimated with an adaptive Markov Chain Monte Carlo (MCMC) algorithm (see Andrieu, De Freitas, Doucet, & Jordan, 2003) in which three separate chains were run (based on different starting values) and convergence was checked after every 200 samples (Gelman & Rubin, 1992). For the group- and subject-based models, the respective numbers of post-convergence samples collected were 4500 and 6000. Model comparison was based on Bayesian Information Criterion (BIC; Schwarz, 1978). Bayes factors (BFs) were computed with a set of MATLAB functions (Schwarzkopf, 2015), using a default scaling factor (*r*) of .707 (see Rouder, Speckman, Sun, Morey, & Iverson, 2009; Morey, Rouder, & Jamil, 2018). All BFs are reported as BF10, where values >1 indicate evidence favoring the alternative hypothesis, and positive values <1 indicate evidence favoring the null hypothesis.

2.2 RESULTS AND DISCUSSION

During the encoding phase, the mean response time (RT) for making pleasantness judgments was 1424 ms (*SD* = 246). For the test phase, the mean

RTs for the temporal and confidence judgments were 2720 (*SD* = 794) and 1064 ms (*SD* = 386), respectively. Figure 2A displays the group-level error distribution of temporal responses. For completeness, the scatterplot between the correct and responded locations across trials is reported for this and all other experiments in the Appendix (Figures A1-4). Centered on each study trial, the mean absolute error was 7.69 trials (*SD* = 13.51). To further parameterize temporal precision, the error distribution was modeled by a normal function with a mean of zero and the standard deviation corresponding to precision error (hereafter, σ). The resulting maximum a posterior probability (MAP) estimate of σ was 75.32 (95% highest density interval [HDI] = [73.93, 76.80]). Modeling individual subjects gave rise to a similar level of precision error: *M* = 74.95 and *SD* = 8.32 (see Figure 2B).

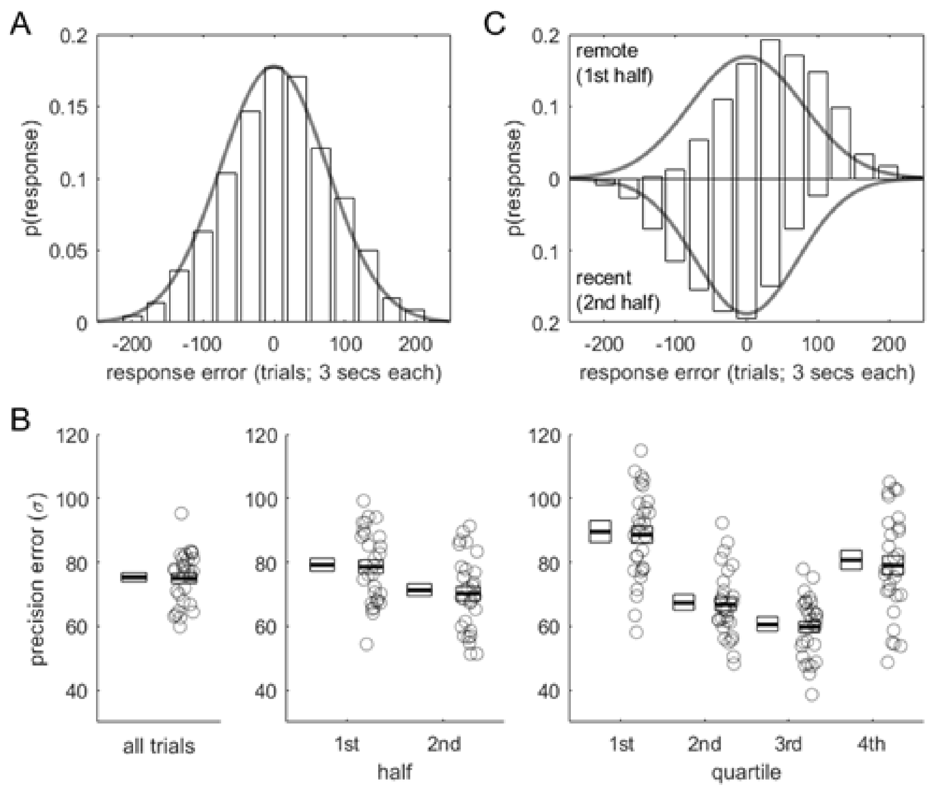

Figure 2. Behavioral Results for Temporal Judgments in Experiment 1
(A) Histogram of the group-level distribution of temporal response errors, centered (at 0) on each study trial. Positive/negative values respectively reflect responses that are more recent/remote than the correct position. The line corresponds to the estimated normal function of the 1-parameter (precision-only) model: $\mu = 0$ and $\sigma = 75.32$.
(B) The precision (σ) results for all, halved, and quartiled trials. For each pair of box plots, the left one indicates the MAP estimate and 95% HDI for the group analysis, and the right one indicates the mean±SEM for the subject-wise analysis. The results for individual subjects are plotted with circles.
(C) Histograms of the group-level distributions of temporal errors, separated according to the first (remote) and second (recent) halves of the study list. The line plots indicate the best-fitting normal functions for each condition ($\mu = 0$; remote $\sigma = 79.19$; recent $\sigma = 71.23$). Data for the recent half is arbitrarily plotted downward for clarity.

2.2.1 Changes in Temporal Precision Across Time

Next, we investigated whether temporal precision changes with passing time by dividing the encoding list in half and applying the modeling procedure described above to the error distribution for each half. As displayed in Figures 2B and 2C, group-level precision error was higher for the first half (hereafter, *remote*; σ = 79.19 [77.24, 81.28]) than the second half (*recent*; σ = 71.23 [69.48, 73.22]). Subject-wise modeling confirmed this difference was significantly different (remote: M = 78.59, SD = 11.38; recent: M = 70.21, SD = 11.70), $t(28) = 2.84$, $p = .008$, $BF_{10} = 3.78$.

Closer examination of the temporal response distributions revealed that precision error was not only lower for recent compared to remote trials, but was considerably lower for the middle of the encoding list. Figure 2B displays this pattern for trials binned into quartiles. To better isolate the remote-recent effect from the difference in precision for the middle versus extreme quartiles, the subject-wise σ estimates of the quartiled data were submitted to a 2 (remote, recent) × 2 (middle, end) ANOVA. The analysis revealed main effects of both factors (respectively, $F_{1,28}$ = 8.72 and 227.56, $p = .006$ and $p < .001$) but no interaction ($F < 1$). Planned contrasts confirmed the remote-recent difference in precision for both the first versus fourth quartiles (M = 88.60 and 79.02, SD = 14.17 and 16.15), $t(28) = 2.20$, $p = .036$, $BF_{10} = 1.57$, and the second versus third quartiles (M = 66.78 and 59.75, SD = 10.63 and 9.50), $t(28) = 3.64$, $p = .001$, $BF_{10} = 30.82$.

2.2.2 Response Bias in Temporal Judgments

By comparing the temporal error distributions to the corresponding model fits (see Figures 2A and 2C), it becomes apparent that errors overall tended to be more recent than remote. We therefore tested this effect by adding a bias parameter (μ), corresponding to a shift in the mean of the normal distribution, to the modeling procedure. The resulting fit of the 2-parameter (μ, σ) model, alongside that of the σ-only model, is shown in Figure 3A. The group-based μ estimate of 7.48 [5.62, 9.16] was consistent with the subject-based results (*M* = 7.70, *SD* = 13.49; see Figure 3B), which were statistically greater than zero, $t(28) = 3.07$, $p = .004$, $BF_{10} = 8.72$. Moreover, the estimates of σ for the 2-parameter model resembled those of the original model (group: 74.96, [73.69, 76.32]; subjects: *M* = 73.34, *SD* = 8.33). Highlighting the importance of bias, despite it adding complexity, model comparison indicated that including it was preferable to excluding it (*BIC* = 63,502.01 and 63,457.38, respectively).

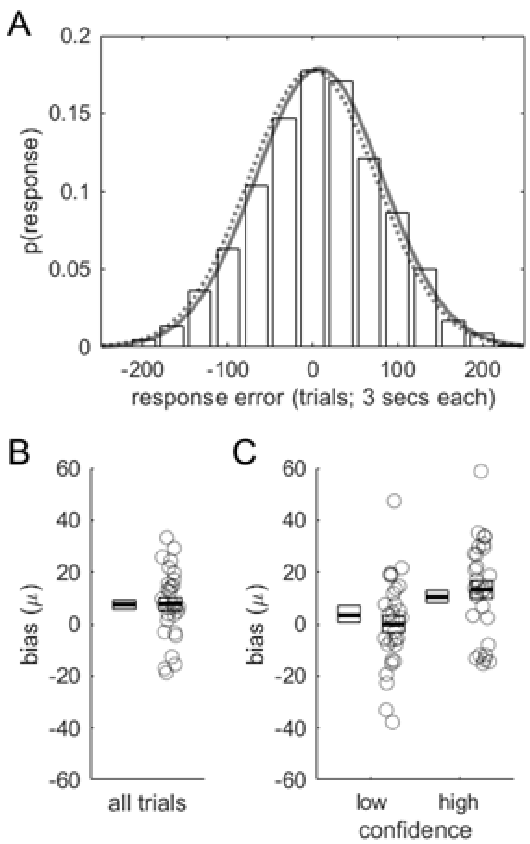

Figure 3. Effects of Response Bias on Temporal Judgments in Experiment 1

(A) Group-level distribution and 1-parameter model fit (dotted line) of temporal response errors, as shown in Figure 1A, along with the fit of the 2-parameter model (solid line)—precision (σ = 74.96) and bias (μ = 7.48).
(B) Box plots indicating the group-wise (MAP, HDI) and subject-wise (M±SEM) results for the bias parameter, with data for individual subjects also plotted.
(C) Bias parameter results corresponding to separate modeling of trials in which low vs. high confidence was designated (with a 50% cutoff).

2.2.3 Guessing in Temporal Judgments

In keeping with prior studies that have often used continuous response paradigms to dissociate precision from guessing (e.g., Brady et al., 2013; Harlow & Yonelinas, 2016; Zhang & Luck, 2008), we next fit the temporal error distribution with a 3-parameter model that included σ and μ (as above) and parameterized the proportion of guessing trials (hereafter, λ) with a uniform distribution. The estimates for precision error and bias were similar to those reported above: σ = 74.81 [73.33, 76.49] (*M* = 72.69, *SD* = 8.16) and μ = 7.68 [5.37, 9.43] (*M* = 7.82, *SD* = 13.57). Guessing, however, was near zero for both group- (5.0e-5 [5.0e-5, .009]) and subject-based models (*M* = .0075, *SD* = .016). To further test that guessing was negligible, we directly compared the 2- and 3-parameter models, the results of which favored the former (respectively, *BIC* = 63,457.38 and 63,467.84).

To ensure that any subtle differences in guessing across the encoding list were not missed, we additionally included it in modeling the data binned into halves. Notably, we excluded the bias parameter for these analyses, as preliminary analyses showed that it varied widely across halves of the encoding list (likely due to the boundaries of the list). The λ estimates were again near zero for both the first (1.6e-4 [1.7e-4, .014]) and second halves (3.6e-4 [4.2e-4, .015]), with the subject-wise means suggesting that there is no strong evidence for guessing rate changes across time (respectively, *M* = .0088 and .014, *SD* = .020 and .030), $t(28) = .73$, $p = .47$, $BF_{10} = 0.73$.

2.2.4 Confidence Judgments

Although confidence judgments were included in the design to follow previous studies (e.g., Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016), we suspected from the outset that subjects might be unable to treat them distinctly from temporal judgments, and we thus had no specific hypotheses regarding differences in confidence. For completeness, though, we summarize the confidence results and then consider them in the context of the modeling approach described above.

Confidence judgments were spread considerably across the 0-100% scale (M = 53.44, *SD* = 9.74). Contrary to what might be expected if confidence tracked temporal precision, it was not correlated with the absolute value of error (r = .0088, p = .51; cf. Harlow & Yonelinas, 2016). Nonetheless, the ratings were binned into high and low confidence (using a 50% cutoff), and the 3-parameter (σ, μ, and λ) modeling procedure was applied separately to the data from each bin. Across high versus low confidence, no decisive evidence of any differences emerged for precision (respectively, σ = 75.75 [73.76, 77.600] and 74.02 [71.62, 76.24]; M = 71.12 and 71.52, SD = 10.09 and 11.42), $t(28)$ = .18, p = .86, BF_{10} = 0.67, or guessing (λ = 6.6e-6 [9.8e-5, .013] and 3.9e-4 [1.8e-4, .015]; M = .020 and .016, SD = .047 and .044), $t(28)$ = .27, p = .79, BF_{10} = 0.67. However, bias was higher—such that recent responses were favored more—for high compared to low confidence (respectively, μ = 10.41 [8.17, 12.81] and 3.39 [.68, 7.09]; M = 13.24 and -.016, SD = 18.70 and 17.77), $t(28)$ = 3.73, p < .001, BF_{10} = 16.12.

3. EXPERIMENT 2

As Experiment 1 novelly extended the traditional continuous report task and mixture-modeling approach to the domain of temporal judgments in episodic retrieval, we next attempted to provide a conceptual replication of its main findings. To this end, Experiment 2 involved having subjects again encode a continuous series of pictures and then make successive responses about each picture's temporal position and the corresponding level of confidence in that judgment. Additionally, there were two noteworthy departures from the previous experiment — one related to a design modification and the other related to an advance in the analysis approach — which we describe in turn below.

In Experiment 1, the pictures that served as stimuli were randomly selected from a pool that contained unequal numbers of pictures from animate versus inanimate categories. Prior research has demonstrated, however, that animate stimuli are often remembered better than inanimate stimuli, in line with the idea that there is an adaptive aspect to human memory in which survival-relevant stimuli have an advantage (Laurino & Kaczer, 2019; Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013; Nairne, VanArsdall, & Cogdill, 2017; VanArsdall, Nairne, Pandeirada, & Blunt, 2013; VanArsdall, Nairne, Pandeirada, & Cogdill, 2015). In addition to this difference potentially affecting memory, it is possible that the unequal trial numbers could have caused some stimuli to appear more salient, thereby interfering with our estimation of retrieval precision. Experiment 2 therefore equated the numbers of pictures from the two categories, allowing for an extension of the findings on adaptive memory to our measure of temporal precision, with the straightforward prediction that animate stimuli would be remembered with greater precision.

One consequence of equating trials on animacy was that the length of the encoding phase had to be reduced (from 300 trials in Experiment 1 to 140 trials here) due to the limited size of the stimulus pool. To accommodate having fewer observations, trial-level multilevel modeling was employed to capitalize on the continuous nature of the predictor corresponding to recency (i.e. encoding position), while also being statistically more powerful than the trial binning procedure used in Experiment 1.

3.1 METHOD

3.1.1 Subjects

Twenty-two MU students (13 females, 9 males), 18-22 years of age (M = 19.1, SD = 1.15), participated for course credit. Inclusion criteria for this experiment were the same as in Experiment 1, with the additional criterion that subjects did not participate in that experiment. Written informed consent was obtained in accordance with the MU Institutional Review Board.

3.1.2 Stimuli and Procedure

The stimuli and procedure for Experiment 2 were largely the same as Experiment 1, with the critical exception of the number and type of stimuli used. Serving two purposes, a subset of stimuli (140) from the original pool of 300 pictures were chosen both to shorten the overall encoding duration and to match the number of pictures depicting animate and inanimate objects (as determined by the author). With this smaller pool of 140 pictures, the study duration was 8.2 minutes. A continuous timeline response scale with the same dimensions as that in Experiment 1 was used. However, the labels on the scale only ranged from 1 to 140, with tick marks placed corresponding to every 23 trials.

3.1.3 Data Analysis

Statistical analyses were conducted in R (v.3.5.1; R Core Team, 2018) using the *lme4,* and *brms* packages (Bates, Maechler, Bolker, & Walker, 2015; Burkner, 2017). To assess evidence for predictor variables in the multilevel analyses, a model comparison approach was used. Omnibus effects were first contrasted with an intercept only model, based on comparing the Bayes Factors (*BFs*) calculated with a bridge sampling technique (*bridgesampling* package; Gronau & Singmann, 2017). All models were fit using random intercepts for individual subjects and weakly informative priors [$\beta \sim N(0,1)$; $\sigma \sim Cauchy(0,2)$]. For each of four separate chains, 2000 samples were estimated (the first 1000 being warm-up samples). Following guidelines from Rouder, Engelhardt, McCabe, and Morey (2016), ratios of *BFs* were taken to assess evidence for main effects. Statistical interpretations were based solely on Bayesian estimates, although frequentist statistics are also reported (Valentine, Buchanan, Scofield, & Beauchamp, 2019).

3.2 RESULTS AND DISCUSSION

At study, subjects failed to respond on average to 2.24% of trials (around 3 out of 140 trials). As in Experiment 1, study items with missing responses were not analyzed during the test phase. The overall mean study RT was 1421 ms (*SD* = 525). There was no difference in the average study RTs between animate (*M* = 1460 ms, *SD* = 521) and inanimate (*M* = 1394, *SD* = 526) trials, $t(1853) = 0.03$, $p = .973$.

Test items that occurred in the first and last 25 study positions were excluded from analysis. Additionally, test trials for which subjects took more than 10 seconds to make the temporal judgment were omitted. The average RT for the remaining temporal judgments was 3129 ms (*SD* = 1728). Again, there were no differences in test RTs between the animate (*M* = 3289 ms, *SD* = 1816) and inanimate (*M* = 3015 ms, *SD* = 1655) categories, $t(1892.16) = 0.08$, $p = .939$.

Temporal judgments were widely distributed across most of the study period, with a clear reduction in the frequency with which subjects respond near the ends of the scale. Figure 4A shows a histogram of judgments at each study position, whereas Figure 4B displays the distribution of confidence ratings. The majority of temporal judgment trials were associated with higher confidence ratings (*M* = 59.80, *SD* = 23.63), as evident by the distribution shift in Figure 4B above 50%. There was no difference in confidence ratings between animate (*M* = 60.31, *SD* = 25.86) and inanimate (*M* = 59.44, *SD* = 21.93) trials, $t(1892.5) = 0.85$, $p = .395$. Figure 5A displays the distribution of errors in temporal judgment, showing a distribution similar to Experiment 1. On average, the error between where subjects responded and the correct study position was 27.9 items (*SD* = 20.6).

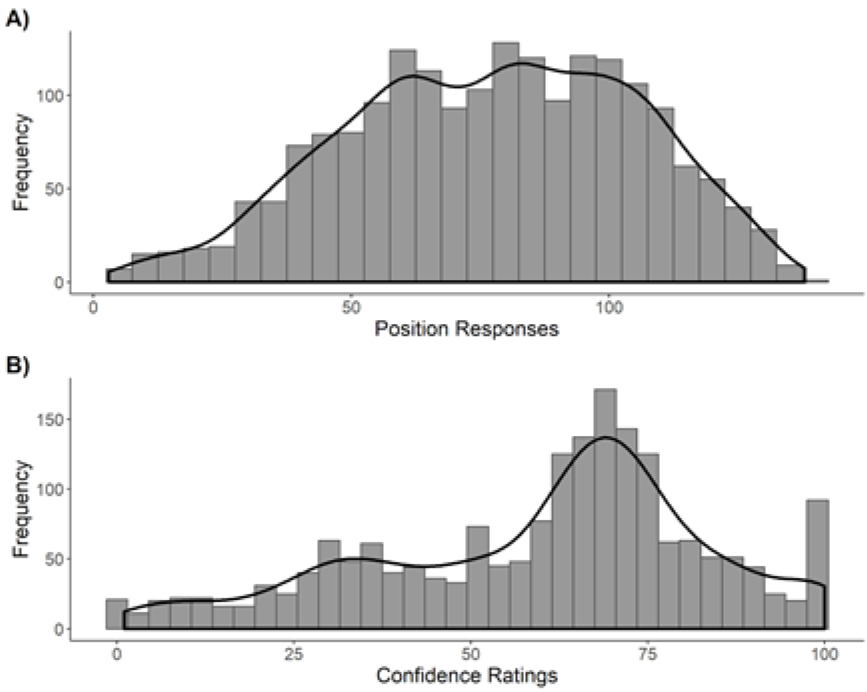

Figure 4. Summary of Temporal and Confidence Judgments in Experiment 2

(A) Histogram of temporal judgments, collapsed over the whole group of subjects, at each position along the timeline representing the study phase.

(B) Histogram of the group-level confidence ratings at test, indicating a general bias to respond with higher confidence (e.g., > 50%).

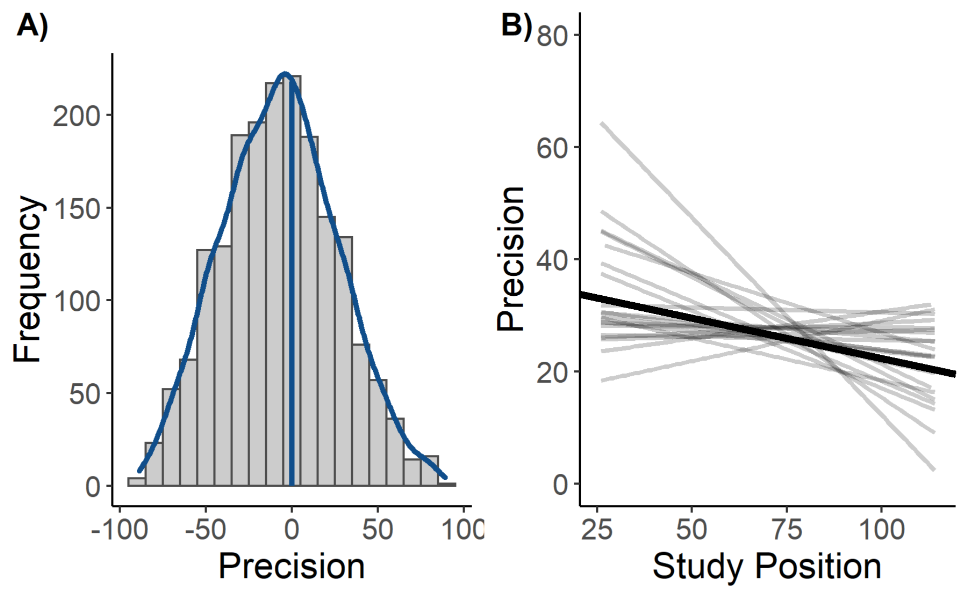

Figure 5. Behavioral Results for Temporal Judgments in Experiment 2
(A) Distribution of errors in temporal judgments across all subjects, with zero indicating responding at the precise location of a previously studied item.
(B) Regression lines showing the negative relationship between study position and the precision of temporal judgments. The group-level line is displayed in black, and the lines for individual subjects are gray.

3.2.1 Standard Mixture-Modeling

The standard mixture-modeling approach used in Experiment 1 was first applied to the current data to assess the replication of the primary effects found in Experiment 1 — namely the negligible guessing rate and the response bias in temporal responses. Using a model comparison approach, the 1-parameter normal distribution mixture model (without any additional parameters) was first fit (σ = 34.68, 95% HDI [33.54, 35.83]) and compared to a 2-parameter model (σ = 33.91 [32.93, 35.06]) that included the recency bias parameter (μ = 7.10 [5.45, 8.72]). The 2-parameter model had a lower *BIC* (18815.54) than the 1-parameter model (*BIC* = 18886.61), indicating that including the recency bias term was preferred.

Next, to examine the role of guessing with this shortened task, the 2-parameter (σ, μ) model was compared to a 3-parameter model (σ, μ, λ). The resulting MAP estimates for the 3-parameter model were as follows: σ = 33.98 [33.03, 35.23]; μ = 7.41 [5.73, 8.73]); and λ < 0.001 [0.000, 0.002]. The 2-parameter model again gave rise to a lower *BIC* (18815.54) than the full model (*BIC* = 18824.93). These modeling results thus replicate Experiment 1 by demonstrating a recency bias in temporal judgments as well as negligible guessing.

To assess how model estimates varied as a function of recency, we divided the encoding list into recent and remote halves. Based on modeling the group-level data, precision was similar across the recent (27.68 [26.63, 28.69]) and remote halves (27.77 [26.52, 28.86]). When repeating this procedure for each subject, a similar pattern emerged. That is, there was no significant difference in precision between the recent (M = 26.42, SD = 4.17) and remote halves (M = 26.78, SD = 4.34), $t(20) = 0.54$, $p = .592$. The corresponding bias and guessing results are provided in Figure 6. Interestingly, we observed that the recent half data elicited a higher bias parameter (8.17 [8.04, 8.35]) than that for the remote data (7.42 [7.05, 7.40]). This pattern was confirmed via subject-level modeling (recent: M = 8.15, SD = 0.89; remote: M = 7.24, SD = 0.70; $t(20) = 5.27$, $p < .001$). Study half, here used as a rough proxy for elapsed time, therefore appears to influence the degree that subjects tend to overestimate the recency of an item. Lastly, guessing remained at near-zero levels across the study halves (recent: <0.001

[0.00, 0.004], *M* = 0.001, *SD* = 0.001; remote: <0.001 [0.00, 0.003], *M* = 0.001, *SD* = 0.001), $t(20) = 0.37$, $p = .715$.

As this experiment additionally equated the study items according to animacy category, we also tested for differences in the mixture-modeling parameters across this factor. Specifically, although precision estimates based on the group data were lower for animate (28.99 [27.86, 30.75]) versus inanimate stimuli (27.07 [25.73, 28.36]), the subject-level data gave rise to no significant difference between the two categories (animate: *M* = 27.33, *SD* = 4.85; inanimate: *M* = 26.55, *SD* = 4.20), $t(18) = 0.98$, $p = .340$. Similarly, the bias parameter was also found to differ across categories based on the group data (animate: 7.58 [7.38, 7.76]; inanimate: 7.79 [7.67, 7.95]) but not based on the subject-wise data (animate: *M* = 7.41, *SD* = 0.93; inanimate: *M* = 7.77, *SD* = 0.79), $t(18) = 1.25$, $p = .229$. Finally, guessing remained near zero in all cases (animate: <0.001 [0.00, 0.004], *M* = 0.001, *SD* = 0.001; inanimate: <0.001 [0.00, 0.002], *M* = 0.001, *SD* = 0.001; $t(18) = 0.93$, $p = .366$). The bias and guessing estimates according to animacy are provided in Figure 6 for completeness.

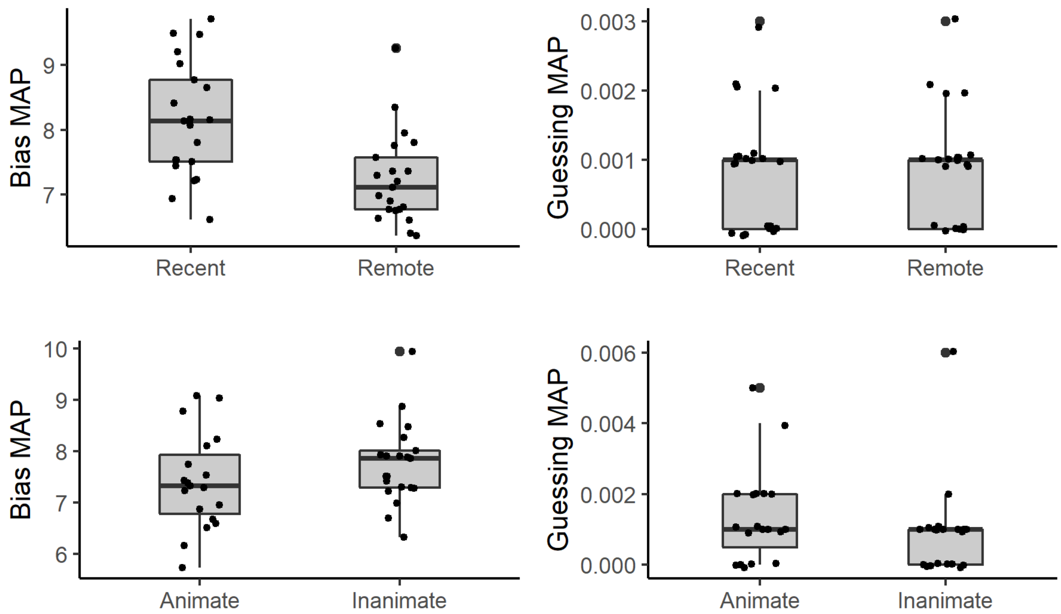

Figure 6. Mixture-Modeling Results for Experiment 2
The left and right columns provide estimates of response bias (μ) and guessing (λ), respectively. The top row provides estimates from mixture-models that were applied separately to the first (remote) and second (recent) halves of the study list. The bottom row provides modeling results separated for the animate and inanimate categories. The box plots display the median and quartile values, along with individual subject results shown by the dots.

3.2.2 Trial-level Modeling of Recency Effects

The foregoing analysis strategy, as well as that of Experiment 1, assessed recency effects in temporal precision via binning the encoding position into discrete halves and quartiles. While this approach revealed consistent precision estimates both at the subject and group level, it is also advantageous to analyze precision differences using the continuous variable of encoding position in its natural form. To do this, trial-level response errors — i.e. the actual temporal position of a trial along the timeline minus the response location — were analyzed using Bayesian multilevel regression models (trials nested within subjects), allowing us to track changes in precision across the entire encoding period instead of relying on binning into recent and remote periods.

Our main prediction for the trial-level analyses was that the basic recency effects found in Experiment 1 would be replicated. That is, we would observe evidence in favor of an encoding period effect where more recent items are associated with better temporal precision. To assess this, four separate models were fit. The first model included all relevant predictors (study position, confidence rating, animacy), whereas the remaining three models each excluded one of the predictor variables. *BFs* were calculated as the ratio of evidence for the addition of a certain predictor, controlling for the effects of the other (e.g., the comparison of the full model which includes study position to the model with every other predictor except study position). Table 1 shows parameter estimates for all models. Replicating Experiment 1, but contrary to the binned results reported above for the current experiment, there was strong evidence for an effect of study position, BF_{10} = 1.66e^{+12}, $F(1,1873.59) = 65.57$, $p < .001$. Specifically, every increase in study position (i.e., more recent) was associated with an improvement (i.e. decrease) in precision of 0.14 items. Figure 5B displays the negative relationship between precision and study position based on the group- and subject-level analyses.

Table 1. Summary of Trial-Level Modeling Estimates for Experiment 2
For each predictor (study position, confidence, animacy), slope estimates, error, 95% Highest Density Intervals (HDI), and Bayes Factors (BF) are presented.

Predictor	Estimate	Est. Error	95% HDI	BF_{10}
Study Position	**-0.14**	**1.91**	**[-0.18, -0.11]**	**$1.66e^{+12}$**
Confidence	0.01	0.02	[-0.03, 0.05]	0.02
Animacy	0.28	0.02	[-1.08, 1.65]	0.73

We next applied the trial-level modeling approach to further test for potential differences in precision according to the animacy category of the stimuli. Figure 7 plots the precision results for each category. As with the standard mixture-modeling results, there was no significant difference in temporal memory precision between animate and inanimate stimuli, BF_{10} = 0.73, $F(1, 1238.86)$ = 0.30, p = .587. From a methodological standpoint, this invariance is useful as it may not be necessary for researchers to control for animacy characteristics while constructing stimuli lists in this type of task, as researchers often need to in terms of frequency of use, imageability, and stimulus meaningfulness.

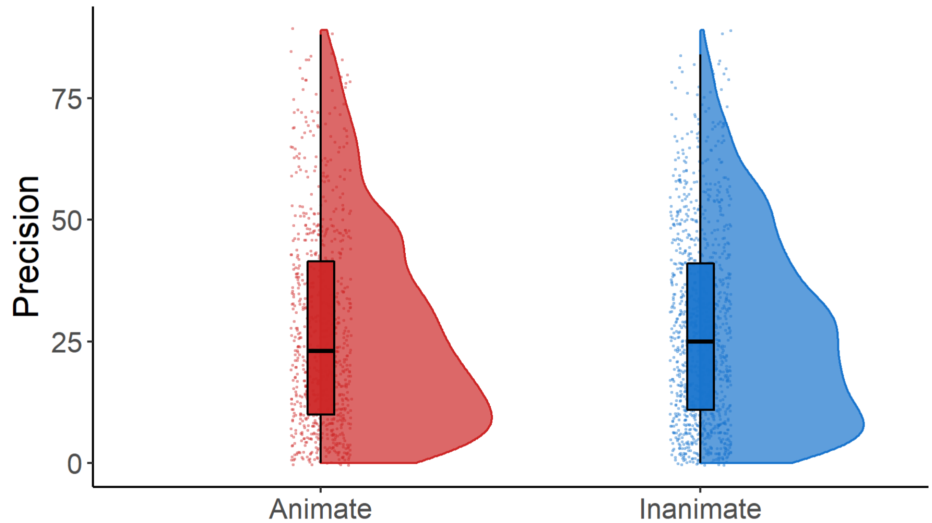

Figure 7. Trial-Level Modeling of Precision According to Animacy in Experiment 2
Group median and quartile box plots of the trial-level estimates in temporal precision, split across animate and inanimate categories. The individual data points, along with the corresponding density plots, are also shown.

4. EXPERIMENT 3

Having established behavioral evidence for the precise retrieval of temporal information in the foregoing experiments, as well as effects of bias and recency on precision, Experiment 3 sought to elucidate the neural correlates of temporal memory precision. As a starting point for hypothesizing about these correlates, we can turn to the extensive neural evidence related to retrieving qualitative (i.e. detailed) episodic information. Such retrieval has been consistently associated with activation in a distributed network of brain regions — sometimes termed the "core recollection network" — including the hippocampus and surrounding medial temporal lobe cortex, medial and lateral parietal cortex, and medial prefrontal cortex (Ekstrom & Yonelinas, 2020; Johnson & Rugg, 2007; Kim, 2010; Rugg & Vilberg, 2013; Spaniol et al., 2009). More recently, these regions have been further divided into separate posterior medial (PM) and anterior temporal (AT) subsystems, comprising what is known as the PMAT framework (Ranganath & Ritchey, 2012; Ritchey, Libby, & Ranganath, 2015). The hippocampus here serves as an integrative hub between the two subnetworks, with the anterior and posterior hippocampus preferentially connected with the AT and PM regions, respectively. The AT subsystem, which consists of the amygdala, anterior ventral temporal cortex, perirhinal cortex, and lateral orbitofrontal cortex, has been observed to be sensitive to changes in general aspects of memory such as familiarity and semantic information. Conversely, activity in the PM subsystem, including the posterior parahippocampal cortex, retrosplenial cortex, angular gyrus, precuneus, and medial prefrontal cortex, correlates with memories that are more episodic-like in nature, accompanied typically by the recollection of details and spatio-temporal context. The precise retrieval of temporal information that is of interest here, therefore, may rely largely on regions within the PM subsystem.

As described in the General Introduction, tests of episodic retrieval typically lack the sensitivity to probe whether general success versus failure is functionally distinct from memory precision. Studies that have employed continuous measures to capture finer fluctuations in the quality of retrieved information have begun to overcome this ambiguity and reveal distinct neural correlates. In one such study, Richter et al. (2016) had subjects retrieve and estimate the location of items positioned along the perimeter of a circle during encoding. The acquisition of fMRI data during a later retrieval phase provided insight into how the neural correlates of retrieval precision could be dissociated from those of other aspects of retrieval, such as success and subjective vividness. Whereas activity in the hippocampus covaried with retrieval success, and precuneus activity correlated with vividness, memory precision was uniquely tracked by a lateral parietal region known as the angular gyrus. In another study, activity in the angular gyrus was again shown to correspond with the precise retrieval of multimodal details, and was dissociated from hippocampal activity (Bonnici, Richter, Yazar, & Simons, 2016). These findings are consistent with prior models of the involvement of cortico-hippocampal networks in memory in which, although the retrieval process might initially rely on hippocampal and medial temporal lobe function (see McClelland, McNaughton, & O'Reilly, 1995; Staresina & Wimber, 2019), it is subsequently supplemented by cortical regions within the broader network (Rugg et al., 2012). Here, the fact that the angular gyrus is part of the PM subsystem is consistent with the idea that judgments of precision are supported by aspects of processing spatio-temporal contextual information.

The notion that retrieval precision is supported by the angular gyrus is also consistent with evidence from event-related potential (ERP) studies that have related episodic retrieval to an effect known as the late parietal effect (LPE; Smith, 1993; Friedman & Johnson, 2000). The LPE is perhaps the most widely studied ERP correlate of memory retrieval, usually occurs in the post-stimulus interval of 500-800 ms, and is topographically maximal over left parietal electrode sites (for a review, see Rugg & Curran, 2007). The successful retrieval of specific details or source information (i.e. recollection) is typically associated with enhanced positivity in this ERP waveform compared to familiarity-based judgments or failed retrieval (i.e. misses). For instance, Wilding (2000) demonstrated that the LPE covaried with the number of correct source judgments associated with a given memory, suggesting that recollection may be better thought of as a graded rather than ‘all-or-none’ process (Rugg, Cox, Doyle, & Wells, 1995). More recently, Murray et al. (2015) used a continuous report task as retrieval and showed that the magnitude of the LPE correlated with precision estimates in a graded manner. Consistent with the idea that the LPE reflects activity in left lateral parietal cortex, Nilakantan et al. (2017) used repetitive transcranial magnetic stimulation (*r*TMS) to suggest a causal role of angular gyrus in retrieval precision. Specifically, targeting the stimulation to this region led to facilitation in the precision of spatial memory estimates as well as a decrease in the LPE, compared to sham stimulation.

In addition to studies that have employed traditional ERP measures, several others have turned to analysis of the rhythmic oscillatory activity present in EEG data to understand episodic memory retrieval. In particular, changes in activity power in the alpha (8-12 Hz) and theta (4-8 Hz) frequency bands has been previously associated with both the maintenance of items in working memory (e.g., Foster, Sutterer, Serences, Vogel, & Awh, 2016) and performance on long-term retrieval tasks (e.g., Waldhauser, Braun, & Hanslmayr, 2016; for review, see Nyhus & Curran 2010), although the direction of these effects can depend on methodological factors. Building on these basic findings, Sutterer, Foster, Serences, Vogel, and Awh (2019) recently used the circular variant of the continuous report task to investigate spatial memory and demonstrated that alpha activity corresponded to the precision of retrieval. This finding suggests that, in addition to assessing ERP correlates, activity in the alpha frequency band, and perhaps the theta band, could provide further insight into the neural processes involved in retrieving precise information from episodic memory.

Together, the EEG/ERP studies described here — although employing a variety of experimental procedures and neural analyses — serve as a starting point for further investigating the electrophysiological correlates of retrieval precision as they relate particularly to temporal information. The purpose of Experiment 3 was thus to extend our novel temporal judgment paradigm to this domain. Additionally, to more robustly test the influence of guessing and bias on temporal judgments, the length of the encoding list in the current experiment was increased considerably. Doing so was intended to allow for truncating trials at the ends of the list, while still including enough of a range of the timeline so as to minimally affect responding. One consequence of considerably increasing the number of trials is the corresponding increase in the duration of the experimental session for each subject. Given that our modeling approach can be applied to the group and subject-level data, increasing the number of trials per subject provides an enhancement in statistical power that allows us to test a smaller sample of subjects (Rouder & Haaf, 2018; Smith & Little, 2018), and is additionally cohesive to an individualized precision-based neuroscience approach (see Gordon et al., 2017).

4.1 METHOD

4.1.1 Subjects

Seven MU students participated in this experiment (six students participated voluntarily and one subject participated for course credit). Inclusion criteria for this experiment included being at least 18 years of age, right-handed, native-English speakers who had normal or corrected vision, and no history of neurological disease. Written informed consent was obtained in accordance with the MU Institutional Review Board. The data from one subject was excluded from analysis due to a technical error that occurred during the experiment. The final sample of six subjects (3 females, 3 males) were 19 to 29 years of age (*M* = 23.83, *SD* = 3.92).

4.1.2 Stimuli and Procedure

As overall memory performance was expected to be poorer than in Experiments 1 and 2 due to the considerably increase in number of study trials, a different stimulus pool, the Large-scale Image Memorability (LaMem) dataset, was used (Khosla, Raju, Torralba, & Oliva, 2015). Originally containing 60,000 diverse images, the LaMem dataset was created with the intention of understanding and predicting visual memorability. To mitigate potential low memory performance, a subset of images (15,794) with memorability ratings above 0.8 (with a rating of 0.8 indicating that 80 percent of subjects remembered that respective item) were used. Each colored picture, randomly selected from the larger subset of images, was presented centrally on a light gray (~70% white) background. Stimuli were displayed on a 24-inch widescreen LCD monitor (1024×768 resolution) viewed at an approximate distance of 1 meter. Stimulus presentation was controlled using PsychoPy3 (Peirce, 2007).

The experimental session consisted of an encoding phase (~30 minutes) followed by a test phase (~50 minutes). Instructions and practice on the encoding phase were administered first. This phase comprised a single block of 1,500 pictures, with each image displayed for 1,250 ms with no inter-trial interval. Given the relatively fast trial pace, subjects did not make study judgments on each item. Instead, the study instructions were to pay attention to each image for the entire time it was presented and to try and remember as many details about the image as possible for an upcoming memory task. Subjects were informed of the number of trials and total time of the encoding phase (31.25 minutes) prior to its start.

Following the encoding phase, subjects received instructions and practice on the test phase. A random subset of the study trials (500) selected from evenly dispersed bins across the study period, were presented in a randomized order. Each test trial began with a picture centrally presented for 2,500 ms. The picture then disappeared, and the time scale was presented. The time scale indicated *Trial 1* on its left and *Trial 1,500* on its right, with unlabeled tick marks across the scale. On each trial, the subject used the computer mouse to move a marker across the scale to judge, as precisely as possible, the position at which they thought the picture occurred at study. To finalize their response, subjects used the left mouse button to click the position on the scale and then click a button labelled *accept* located beneath the scale. After subjects responded, a central plus sign was displayed for 500 ms until the next trial began.

4.1.3 EEG Acquisition and Processing

EEG data were acquired during both the study and test phase of the experiment from 59 Ag/AgCl electrodes using a BrainAmp Standard system (Brain Vision LLC; Durham, NC, http://www.brainvision.com). Electrodes were embedded in an elastic cap (Easycap, Herrsching, Germany, http://www.easycap.de) at the following locations of the 10-20 system; Fpz/1/2, AFz/3/4/7/8, Fz/1/2/3/4/5/6/7/8, FC1/2/3/4/5/6, FT7/8, Cz/1/2/3/4/5/6, T7/8, CPz/1/2/3/4/5/6, TP7/8, Pz/1/2/3/4/5/6/7/8, POz/3/4/7/8, and O1/2. Data were acquired in reference to an electrode placed at FCz, and a ground electrode was located at FT10. Two electrodes were placed to the left and right mastoids (M1 and M2) for offline re-referencing. Vertical and horizontal EOG were recorded with electrodes placed on the outer canthi (LO1 and LO2) and below the left eye (IO1). Electrode impedances were adjusted to below 5 kΩ prior to the start of encoding. Data were recorded at a sampling rate of 1 kHz with an amplifier bandwidth of .01-100 Hz.

Offline processing of the EEG data was implemented using the *MNE* module (Gramfort et al., 2014) in Python v. 3.6.4. The continuous data were band-pass filtered from .05 Hz to 50 Hz, down-sampled to 200 Hz, re-referenced to the mastoid average, and then epoched from -200 ms to 2,000 ms relative to stimulus (picture) onset. Independent components analysis (ICA) was used to manually identify and remove artifacts (e.g., due to blinks and eye/head movements) in the data based on scalp topography and spectral composition (Jung et al., 2000). To create the time-frequency representations, EEG decomposition was performed using Morlet wavelets. Specifically, frequency (Hz) estimates were extracted at each electrode × time point for the alpha (8-12 Hz) and theta (3-8 Hz) ranges.

4.1.4 Data Analysis

The behavioral analysis strategy, following Experiment 1, utilized mixture-modeling to assess temporal precision at the group and subject level. Multi-parameter models were tested to examine guessing and recency bias factors. This is an informative test compared to Experiments 1 and 2, considering the mitigated stimuli truncation issue described previously. We predicted that at the individual-level, temporal precision would closely match the group precision estimates. We also predicted that guessing would be negligible and that subjects would still show a recency bias in temporal judgments. Neurally, ERP amplitudes were tested across several bins of lower- to higher-precision, as well as using trial-level regression analyses. In addition to the regression analyses examined in the *a-priori* selected left parietal old/new effect, regression models were additionally run across all electrode × time combinations (binned into 25 ms intervals) to assess if any potential effects exist outside of the LPE in an exploratory fashion. To accomplish this, we adopted a regression heatmap approach. At each electrode × time point combination, a regression model was fitted. From each resulting regression analysis, both the *p* values and slope coefficients were stored, producing a table (e.g., electrodes as rows and time points as columns), or heatmap, of *p* values and slope coefficients, where each cell is the result of a separate regression analysis. The benefit of this exploratory approach is to visually inspect these heatmaps to identify which electrodes and at what time points there appear to be ERP effects relating to behavioral precision. This serves as a critical first step in formulating *a-priori* hypotheses for future studies. In further exploratory analyses, we adopted this multilevel regression strategy to the oscillatory alpha and theta frequency data. No predictions were made as to where or when potential effects might have occurred.

4.2 RESULTS AND DISCUSSION

4.2.1 Behavioral Results

Given that subjects made no responses to items during the encoding phase, the behavioral analyses focused exclusively on the test phase. Test items that came from the first and last 100 positions of the 1,500-item encoding phase were excluded from analysis to mitigate primacy and recency effects as well as any issues with responding near the ends of the temporal scale. Test trials for which RTs were greater than 10 s were also removed (accounting for only 2.5% of trials on average per subject; SD = 2.3). Of the remaining trials, the average RT was 3050 ms (*SD* = 1616; Figure 8B).

Turning to the error associated with making temporal judgments, Figures 8A and 8C respectively display the group- and subject-level error distributions centered on each item's correct position. The average absolute error distance for the group data was 392.7 trials (*SD* = 282.4), with the subject data exhibiting mean values ranging from 344.1 to 437.1 trials. Temporal precision was first modeled by a normal function with a mean of zero and standard deviation representing precision error (σ). The resulting model fits are also displayed in Figure 8. The MAP estimate of σ based on the group data was 483.36 (95% HDI = [471.35, 495.90]). Applying this mixture model to individual subjects revealed a similar level of precision error: M = 483.00, SD = 37.95.

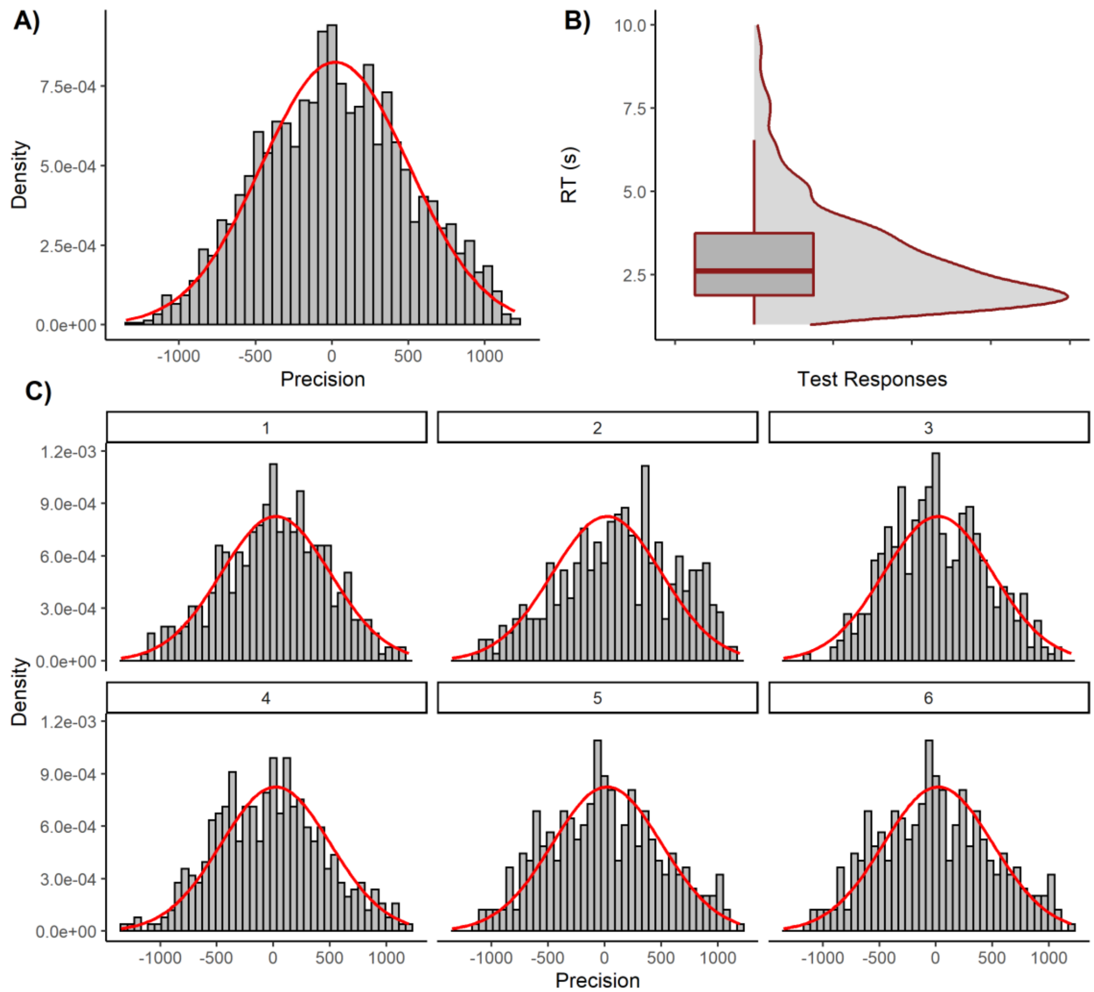

Figure 8. Behavioral Results for Temporal Judgments in Experiment 3
(A) Density plot representing the group-level distribution of error in temporal judgments, along with the best-fitting normal curve (in red).
(B) Box plot of response times during the testing phase, along with a density plot representing the distribution of RTs, showing that the majority of responses were at the lower end around the median value.
C) Density plots representing the distribution of precision error judgments, along with the fitted curve in red for each of the six subjects in the experiment. While the distribution varied slightly from subject to subject, the fitted curves point to a similar normal distributional shape in temporal memory retrieval.

Following the analysis approach of the first two experiments, the encoding list was next divided in half to test for changes in precision with passing time. Applying the same modeling procedure described above to the errors for each half revealed that, based on the group-level data, remote trials (466.50 [449.02, 486.28]) were associated with better precision than recent trials (499.85 [480.88, 520.73]; see Figure 9). When modeling errors at the subject-level across remote (M = 466.28, SD = 42.64) and recent (M = 498.38, SD = 56.16) trials, there was no statistically significant difference in mixture-modeling precision estimates, $t(5) = 1.21$, $p = .281$. However, analyzing these data using the more sensitive technique of trial-level multilevel modeling resulted in a significant difference in precision across the recent and remote halves, $t(2525) = 3.19$, $p = .001$. Importantly, as with the results of standard mixture modeling applied to the group data, this difference was in the opposite direction as that observed for Experiments 1 and 2.

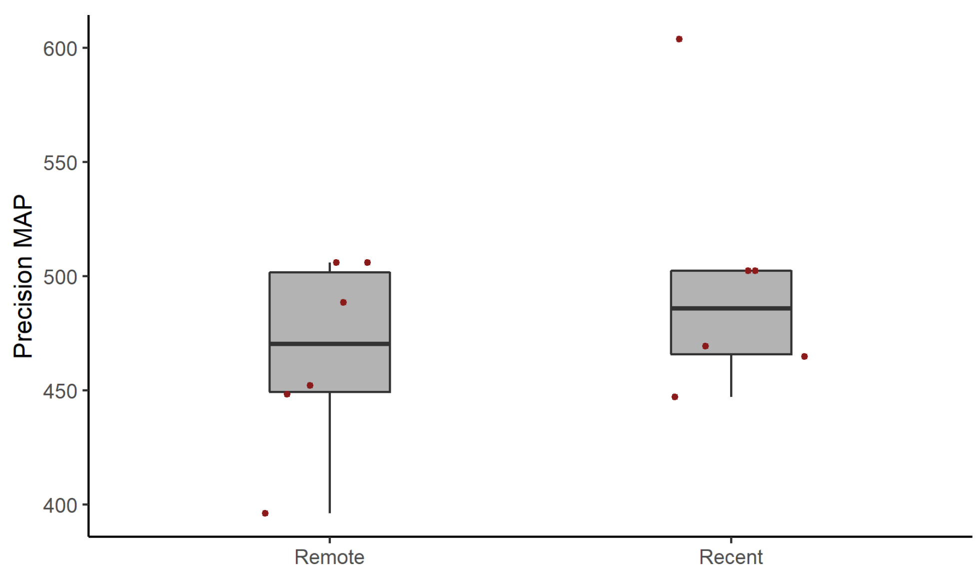

Figure 9. Effects of Recency on Temporal Precision in Experiment 3
Box plots corresponding to the group-level precision estimates for the remote and recent halves of the study phase. Dots indicate each subject's precision estimate.

Based on visual inspection of the overall error distribution in Figure 8A, there was no apparent pattern of response errors showing the recency bias that was evident in the first two experiments. To confirm this statistically, a bias parameter (μ) corresponding to a shift in the overall mean of the fitted normal distribution was added to the modeling procedure. This 2-parameter (μ, σ) model gave rise to a group-based μ estimate of 20.37 [15.9, 37.36]. While the group level estimate suggests a recency bias similar to that in Experiments 1 and 2 (e.g., the highest density interval does not overlap with zero), when considering the subject-based results, average bias estimates were not statistically greater than zero (M = 20.63, SD = 55.87, $t(5) = 0.90$, $p = .204$). Moreover, half (3 out of 6) of the subjects showed a negative bias effect, in the opposite direction of the group level bias estimate, indicating a tendency to respond more remotely. Additionally, using a model comparison approach, the 1-parameter model (σ; *BIC* = 26837.14) was preferred over the 2-parameter model (σ, μ; *BIC* = 26842.21), confirming the subject-based analyses that the recency bias effect found in the first two experiments was not a significant contributor in the current experiment.

Similar to Experiment 2, we also tested for differences in response bias across the two halves of the encoding period. From the group level mixture-modeling, a dichotomy emerged where remote trials were associated with a tendency for subjects to respond that items were more *remote* than what they were (μ = -27.83 [-29.90, -26.36]). Trials from the recent half, however, showed the typical recency bias effect found in Experiments 1 and 2 (μ = 31.06 [29.35, 33.01]). This difference was confirmed at the subject level, with a significant difference evident between remote (*M* = -27.89, *SD* = 5.19) and recent trials (*M* = 31.81, *SD* = 7.64), $t(5) = 49.53$, $p < .001$. Figure 10 shows this opposing relationship between encoding half and response bias estimates, and is likely the reason why bias estimates at the group level (agnostic to encoding half) revealed bias estimates that were not significantly greater than zero, and why model comparison favored the model that excluded the bias parameter. Recency bias estimates were more pronounced with more recent trials in Experiment 2, and combined with the data in this Experiment suggest that response biases in temporal judgment are vulnerable to when an item was studied.

Lastly, the role of guessing was examined by estimating an additional 2-parameter model that included precision and a uniform guessing parameter (λ). The group-level mixture model estimated λ at near zero (<.001 [0.00, 0.03]). Likewise, the subject-based estimates were also near zero (*M* = 0.002, *SD* = 0.004), mirroring the results found in Experiments 1 and 2. Comparing the 1-parameter model (precision only; *BIC* = 26837.14) to this 2-parameter model (including guessing; *BIC* = 26846.81) again revealed a preference for the former. Finally, contrasting the guessing rates across the two halves of the encoding list also did not reveal any differences (Remote: MAP < 0.001 [0.00, 0.004], subject-level *M* < 0.001, *SD* < 0.001; Recent: MAP < 0.001 [0.00, 0.004], subject-level *M* = 0.001, *SD* = 0.001), $t(5) = 1.35$, $p = .235$. The results of this comparison are provided in Figure 10.

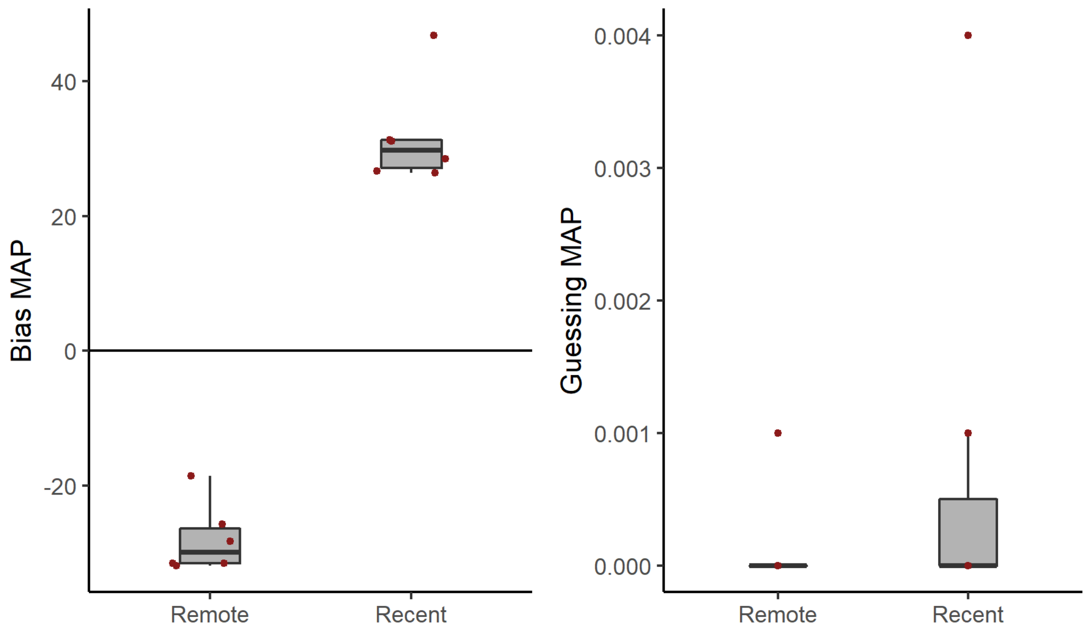

Figure 10. Effects of Response Bias and Guessing in Experiment 3
The box plot on the left displays the group-level estimates of response bias for each half of the encoding phase, whereas the plot on the right provides the corresponding estimates of guessing. Dots indicate the corresponding estimates for each subject.

4.2.2 Event-Related Potentials

Event-related potentials (ERP) were first examined in the context of the left parietal positivity that has been associated with retrieval success and, more specifically, recollection (for reviews, see Friedman & Johnson, 2000; Rugg & Curran, 2007). For this analysis, the ERP amplitudes between 500 and 800 ms after test-item onset were averaged over a group of six left posterior electrodes (CP3/CP1/CPz and P3/P1/Pz) where this effect is typically maximal (e.g., Murray et al., 2015; Rugg & Curran, 2007; also see Scofield et al., 2020). As a first pass, trials were median-split according to better and worse precision (i.e. lower and higher error values, respectively). The resulting ERPs are shown in Figure 11. Using trial-level mixed effects modeling, we did not observe a significant amplitude difference between the better (M = 5.78 μV, SD = 9.12) and worse (M = 5.26 μV, SD = 9.35) precision groups, $F(1, 2861) = 2.52$, $p = .113$. To test for more fine-grained differences, the ERP data were grouped into quartiles, ranging from best (hereafter, Group 1) to worst (Group 4) precision (Group 1: 0-161 trials away from the correct position; Group 2: 162-360; Group 3: 361-609; Group 4: 610-1,448). There was a marginally significant group effect, $F(3, 2859) = 2.56$, $p = .053$. As can be seen in Figure 12, there appeared to be little variation within the first three groups, with a noticeable drop in ERP amplitude with the worst precision group. Pairwise comparisons of the quartiled data revealed no differences among the three groups with the best precision (Group 1: M = 5.88, SD = 9.20; Group 2: M = 5.67, SD = 9.04; Group 3: M = 5.77, SD = 9.20; p's > .05), but the difference between the best and worst (Group 4: M = 4.76, SD = 9.47) quartiles reached significance, $t(2859) = 2.48$, $p = .013$.

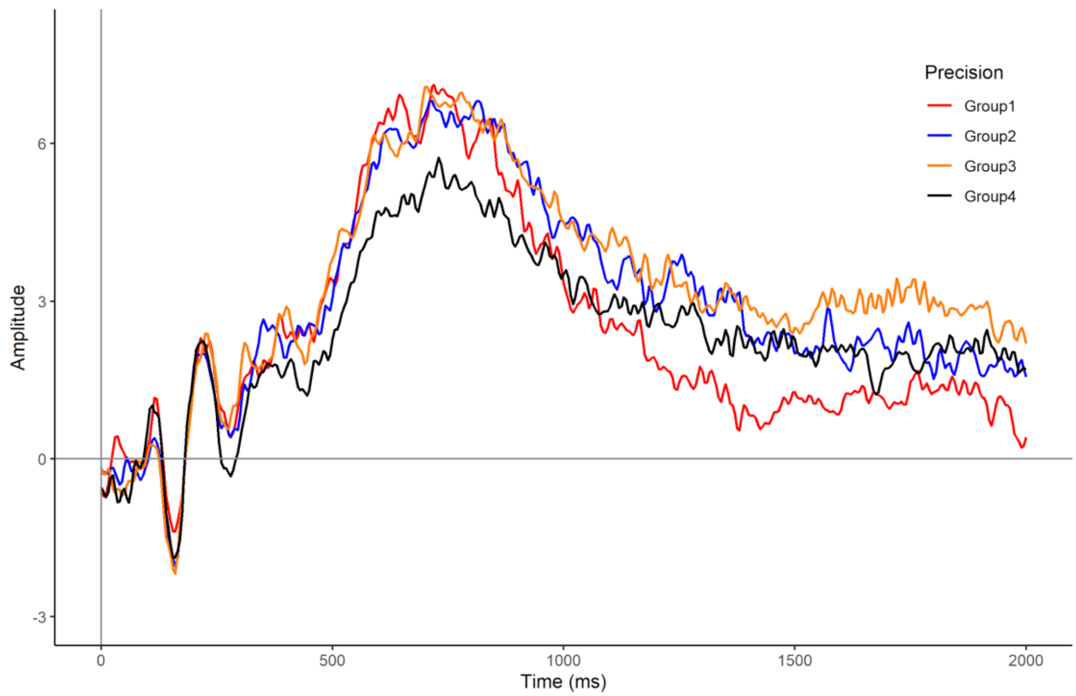

Figure 11. Left Parietal ERPs According to Precision Level in Experiment 3
ERPs averaged over a group of left parietal electrode sites (CP3, CP1, CPz, P3, P1, and Pz) according to four levels of behavioral precision for temporal judgments. Groups 1 through 4 correspond to trials the best through worst precision.

To capitalize on the continuous nature of our precision predictor and as an alternative to binning trials into discrete groups, we next ran a trial-level mixed effects regression model between LPE ERP amplitudes and the continuous precision predictor. The results of this analysis are displayed in Figure 12. Specifically, there was evidence for a significant relationship between ERP amplitude and precision, $b = -0.095$, $t(687.69) = 2.62$, $p = .009$. This result indicated that, as precision values decreased (i.e., corresponding to better precision), ERP amplitudes likewise tended to increase.

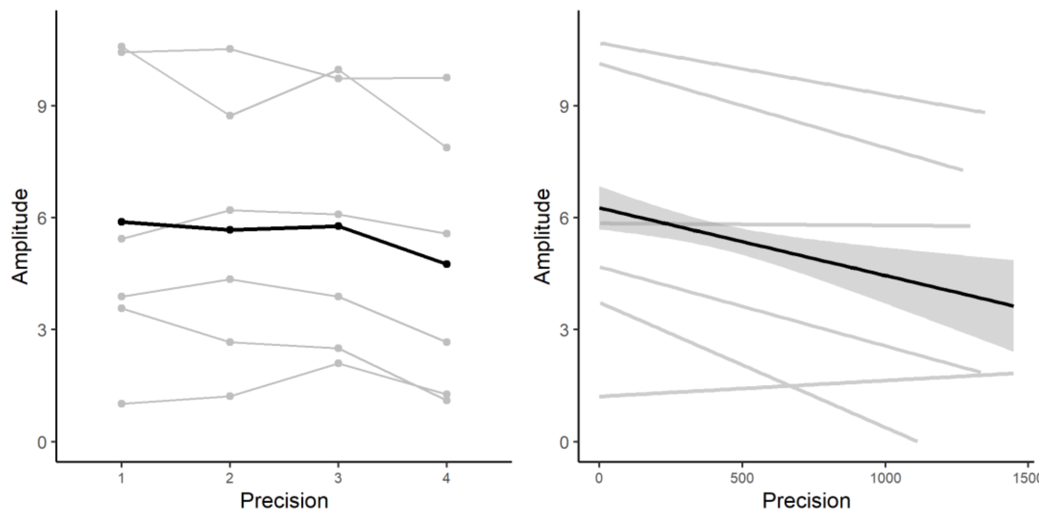

Figure 12. Left Parietal ERP Amplitudes by Precision Level in Experiment 3
The left plot displays average left parietal ERP amplitudes across four binned levels of precision (1 = best; 4 = worst). At right, the regression lines are plotted for the relationship between ERP amplitude and the continuous form of behavioral precision. In both plots, the black and gray lines correspond, respectively, to group and individual subject results.

To widen our scope beyond the *a priori* group of six electrodes and the 500-800 ms time range, mixed effects regression models were estimated in an iterative fashion across all electrode × time point combinations. For each combination, a separate regression model was fitted, from which the betas and *p*-values were stored in matrix form to visually indicate which electrodes and at which time points indicated a precision-amplitude effect. Figure 13 shows the resulting betas and *p*-values as “heatmaps”, where significant effects were evident primarily in about the 600-900 ms latency period, consistent with the timing of the LPE described above. In addition to the significant effects in left parietal electrodes, clusters of significant cells in this 600-900 ms time frame were also apparent in right central electrodes. This may indicate that beyond the more traditionally studied LPE, effects of this nature could be more widespread.

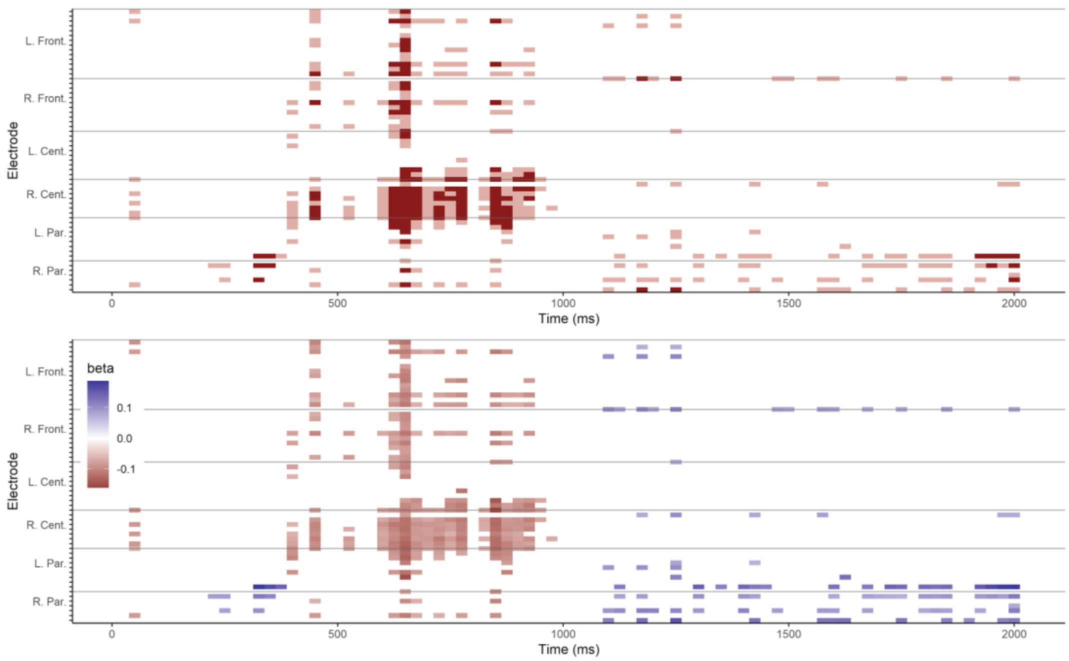

Figure 13. Regression Results of Precision and ERP Amplitude in Experiment 3
Heatmaps displaying the beta (top) and (p-values) from fitting regression models between behavioral precision and ERP amplitude at each electrode and time bin (25-ms) combination.

4.2.3 Oscillatory Activity

In two sets of exploratory analyses, both alpha and theta power were considered in the context of their potential relationship to temporal retrieval precision, as these two EEG frequency bands have been implicated in recognition memory (Herweg, Solomon, & Kahana, 2020; Sutterer et al., 2019; Waldhauser et al., 2016). Average alpha (8-12 Hz) and theta (3-8 Hz) power were computed for each time point and electrode, yielding results in a similar format to those of the foregoing exploratory ERP amplitude analyses (subjects × trials × electrodes × time points). A similar mixed effects multilevel analysis approach was adopted, fitting regressions at each electrode *x* time point combination (nesting trials within subjects and specifying random slopes for the predictors) for each of the two frequency bands.

Figure 14 shows the significance heatmap for the relationship between alpha power and behavioral precision. Averaging across the group of six electrodes associated with the LPE within the 500-800 ms range, there was no significant association between alpha power and memory precision, $b = 0.057$, $t(1.399) = 00.69$, $p = .588$. Upon visual inspection of the significance heatmap, two clusters were visually evident at the $p < .01$ threshold. One cluster (for convenience: right frontal cluster) included electrodes Fpz/2, AFz/4, F1/z/2/4/6, and FC2/4/6 between 950 - 1,100 ms post stimulus onset. Averaging across this cluster, the subject-level regression coefficients can be seen in Figure 15, in which all subjects showed positive slopes above zero. The second cluster (for convenience: left parietal cluster) included electrodes P3/1/z, PO7/3/z, and O1 between 1,175 - 1,400 ms post stimulus onset, in which we did see again positive regression coefficients across all subjects above zero. For both of these observed clusters (right frontal and left parietal cluster), the range of positive slope coefficients indicated higher alpha power were associated with higher (*i.e.* worse) precision.

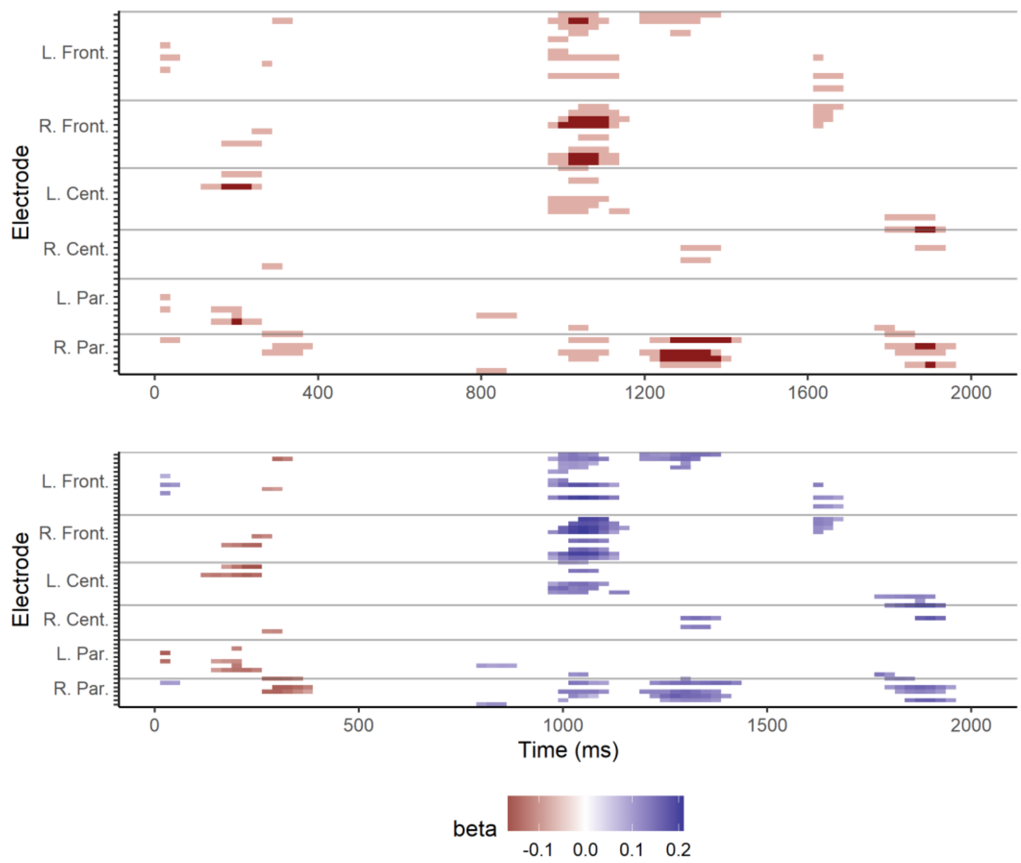

Figure 14. Regression Results of Precision and Frequency Power in the Alpha Band in Experiment 3

Heatmaps displaying the beta (top) and (p-values) from fitting regression models between behavioral precision and alpha (8-12 Hz) power at each electrode and time bin (25-ms) combination.

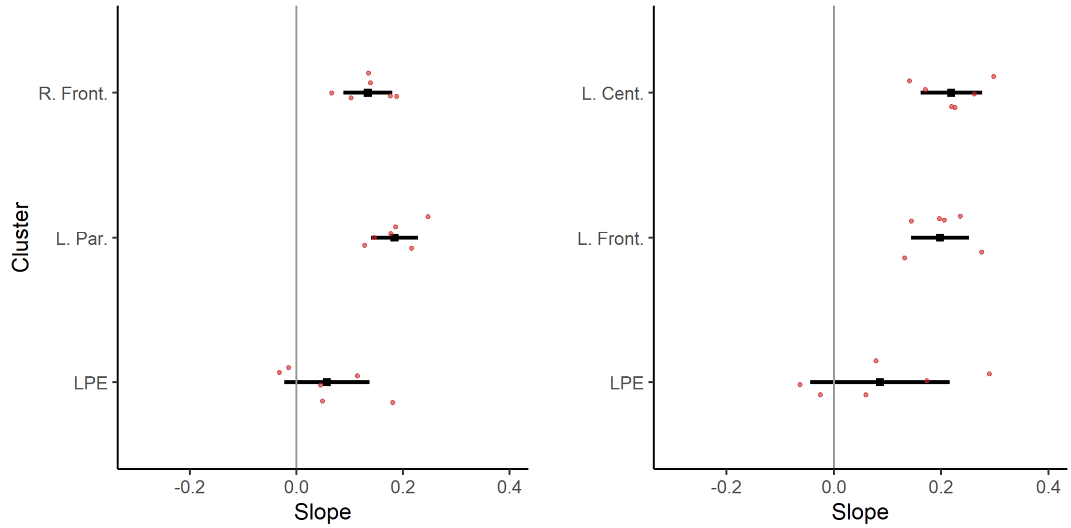

Figure 15. Clustered Regression of Precision and Frequency Power in Experiment 3

Forest plots describe the results of clusters exhibiting a relationship between alpha (left) or theta (right) power and behavioral precision. In each cluster (row), the black square indicates the mean regression coefficient (beta), the error bars correspond to the standard deviation, and the red dots indicate subject-specific regression coefficients.

Figure 16 shows the significance heatmap for the relationship between theta power and behavioral precision. Again, first looking at the 500-800 ms LPE effect, there was no significant precision effect with theta power, b = 0.086, $t(2.59) = 0.89$, p = .449 (Figure 15 shows subject-level regression coefficients overlapping with zero). Upon further visual inspection of the heatmap, two clusters were also evident based on the p <.01 threshold. The first cluster (for convenience: left central cluster) included the T7, C5, TP7, and CP5 electrodes between 925-1,125 ms post-stimulus onset. The left central theta cluster was positively associated with precision, and Figure 15 shows all subjects with positive regression coefficients that do not overlap with zero. The second cluster (for convenience: left frontal cluster) included electrodes Fp1, AF7/3, F7/5, FT7, and FC5 between 1,600-1,700 ms post-stimulus onset, and also showed a positive relationship with precision, as the regression slopes in Figure 15 indicate as they are likewise greater than zero. In a similar fashion to alpha power, increases in theta power were associated with higher/worse behavioral precision.

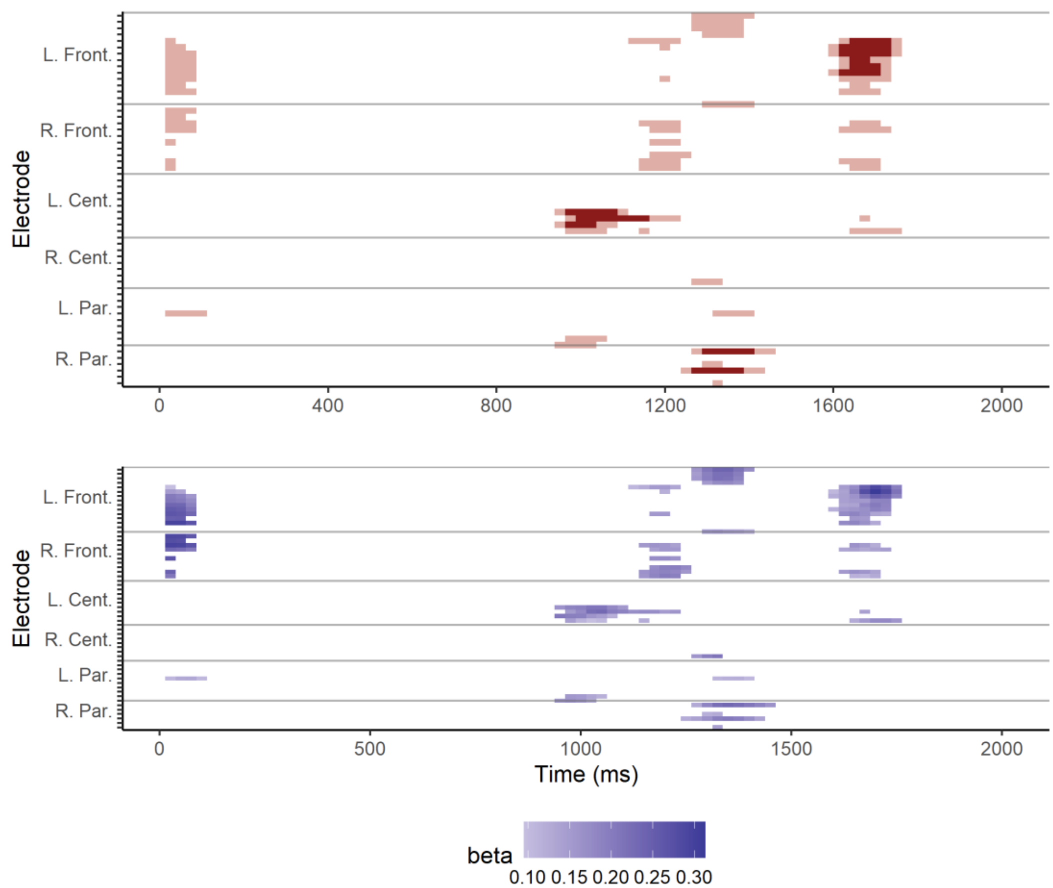

Figure 16. Regression Results of Precision and Frequency Power in the Theta Band in Experiment 3

Heatmaps displaying the beta (top) and (p-values) from fitting regression models between behavioral precision and theta (3-8 Hz) power at each electrode and time bin (25-ms) combination.

5. EXPERIMENT 4

In addition to the novel findings of recency and bias on retrieval precision observed in the three previous experiments, one of the more compelling results is that guessing appeared to play no role in temporal judgments. As studies using continuous report tasks to test spatial memory have consistently indicated the involvement of guessing (e.g., Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Murray et al., 2015), one explanation of the discrepant results pertains to our use of a linear instead of circular response scale. Indeed the choice of a linear scale was intentional, given that the start and end of the encoding list are maximally distinct in time, contrary to how they would be represented if the scale was circular. It is thus unclear whether the characteristics of temporal and spatial retrieval would remain distinct if they are probed with analogous tasks in the same sample of subjects, thereby allowing for direct comparison.

Numerous studies have shown that retrieving spatial and temporal aspects of episodic memory relies on similar neural circuitry within the hippocampus and medial temporal lobe (e.g., Eichenbaum, 2017; Kraus, Robinson, White, Eichenbaum, & Hasselmo, 2013; Salz et al., 2016) that has also been linked to spatial navigation and representations (Herweg et al., 2020; Kjelstrup et al., 2008; Poppenk, Evensmoen, Moscovitch, & Nadel, 2013). The main motivation of this final experiment was to further understand the behavioral components of retrieval precision for both temporal and spatial information and, specifically, to determine the extent to which processes relating to precision and decision-making covary or functionally dissociate across these domains. Whereas the encoding phases of the previous three experiments reported here included no manipulations, as temporal retrieval could be tested with just a standard list of items, the critical manipulation introduced in Experiment 4 was that each picture was presented at a random location on the computer screen during encoding. On a surprise memory test, subjects were then required to judge both *when* and *where* each item was previously presented.

Using a multilevel mixture-modeling approach, performance on the temporal and spatial components of the retrieval task were directly contrasted. In addition, because subjects provided both types of responses on each test trial, our analyses additionally made use of conditionalized data. Our main prediction from these analyses was that there would be significant correspondence, in terms of precision and guessing rates, between the temporal and spatial judgments for a given item. The two judgments were expected to diverge, however, in terms of recency effects. That is, temporal retrieval should be able to rely uniquely on memory strength as a source of information, such that the greater strength of recent items should lead to more precise judgments (Hintzman & Block, 1971). Such strength-based information is uninformative as to the spatial location of the item, given that location was randomly chosen.

5.1 METHOD

5.1.1 Subjects

Two-hundred and one MU undergraduate students were recruited to participate in an online experiment for course credit. Of the original sample, 87 subjects did not complete the experiment and three subjects experienced technological errors, and thus their data were excluded from analysis. The final sample of 111 subjects (85 “female”, 21 “male”, and 5 “other” responses) were between 18 and 51 years of age (*M* = 19.51, *SD* = 5.00). The inclusion criterion for this experiment was only being at least 18 years of age, and informed consent was obtained electronically in accordance with the MU Institutional Review Board.

5.1.2 Stimuli and Procedure

The stimulus pool consisted of 100 color pictures of common, nameable objects, randomly chosen out of a pool of 400 images. These were the same stimuli used in Experiments 1 and 2. Pictures were presented relative to the aspect ratio of each subject’s computer monitor (as subjects were instructed to complete the experiment on either a laptop or desktop computer). For instance, in a standard widescreen monitor with an aspect ratio of 16:10 and a screen resolution of 1650x1050, images were 300 by 300 pixels in size. These presentation characteristics could not be controlled due to subjects completing the experiment on their own device, but the ratios remained constant relative to each screen. Pictures were presented on a light gray (~70% white) background. Stimulus presentation was controlled and behavioral responses were collected using PsychoPy3 (Peirce, 2007), and the experiment was hosted on Pavlovia (pavlovia.org).

The experimental session (~45 minutes) consisted of an encoding phase followed by a test phase. Instructions and practice on encoding were administered first, keeping subjects naïve about the nature of the test phase until immediately prior to its start. The encoding phase comprised a single block of 100 pictures. The images were restricted to be presented within a red bordered rectangle that comprised 92.8% of the screen width and 89.4% of the screen height. In keeping with our example of a 16:10 aspect ratio and screen resolution of 1650x1050, this would imply that images were restricted to a centered rectangular section of the screen with a size of 1531 by 938 pixels. Each encoding trial began with a plus sign presented for 500 ms in a randomized location within the rectangular section on the screen to prepare subjects where the image was presented on the screen. Then, centered at the same location of the plus sign, the image was presented for 3000 ms. Subjects were instructed to rate the pleasantness of each picture on a 4-point scale (very pleasant, somewhat pleasant, somewhat unpleasant, and very unpleasant) via key presses mapped to their left little through index fingers (keys H, J, K, and L, respectively). The next trial then proceeded using a new, random location on the screen. Subjects were informed of the number of trials and the total time of the encoding phase (about 6 minutes) prior to its start.

Following the encoding phase, subjects received instructions and practice on the test phase, which consisted of one single block of 100 trials. All pictures from encoding were presented in a newly randomized order during the test phase. On each trial, subjects made two separate judgments. On each judgment, the studied image was centrally presented for 2500 ms. One of the judgments was the temporal judgment used in the first three experiments. A time scale was presented, ranging from trial 1 on the left to trial 100 on the right, with no label markers between the endpoints. The subject used the computer mouse to click on the scale, as precisely as possible, the position at which they thought the picture occurred at encoding. The second type of judgment was a spatial judgment. On each trial, subjects used the computer mouse to click, as precisely as possible, the spatial location on the screen where they thought the picture occurred at encoding. After subjects provided both retrieval judgments, the next trial began. The order of the judgments was randomized and manipulated between-subjects, with half of subjects providing the temporal judgment first on each trial (n = 52), and the other half providing the spatial judgment first on each trial (n = 59).

5.1.3 Data Analysis

The behavioral analysis strategy was similar to Experiment 2. To assess the relationships between temporal and spatial precision specifically at a trial/item level, and to capture the continuous nature of both judgment predictors, multilevel modeling of the response errors were used instead of the mixture-modeling approach first used in Experiment 1. We predicted that trials in which the temporal context is precisely retrieved would also be accompanied by better spatial precision as well, manifesting as significant correlations/regression coefficients between the two judgments. Additionally, recency effects in both temporal and spatial judgments were examined (e.g., changes in precision across the study period). We predicted that both temporal and spatial judgments would show similar recency effects like those found in Experiments 1 and 2.

5.2 RESULTS AND DISCUSSION

5.2.1 Response Times

During encoding, the pleasantness judgments were made with an average RT of 1.38 s (*SD* = 0.47). At test, subjects gave two separate judgments for each picture. To avoid issues related to responding near the edges of the temporal and spatial scales, the first and last five trials of the 100-item encoding list were trimmed, and the pictures presented along the borders of the screen (5% of the pixels away from each edge) were excluded from analysis, leaving on average 72 viable trials per subject (*SD* = 3 trials). Trials in which subjects did not respond within a predefined 10 s range were excluded. For the remaining trials, subjects responded to the temporal and spatial tasks with respective mean RTs of 2.39 s (*SD* = 1.24) and 1.23 s (*SD* = 0.47), with spatial judgments occurring significantly faster than temporal judgments, $F(1,87) = 238.51$, $p < .001$. The between-subjects order in which subjects completed the judgments interacted with the type of judgment, $F(1,87) = 38.17$, $p < .001$. That is, temporal judgments were significantly slower when they were made first (*M* = 2.81, *SD* = 0.68), compared to when they were made after a spatial judgment (*M* = 1.93, *SD* = 0.55), $t(77.13) = 6.71$, $p < .001$. There was no significant difference in spatial judgment RTs depending on whether it was performed first (*M* = 1.24, *SD* = 0.46) or second (*M* = 1.21, *SD* = 0.35), $t(85.81) = 0.34$, $p = .738$. Figure 17 shows RT values across the two judgment types as well as the order in which subjects provided their judgments.

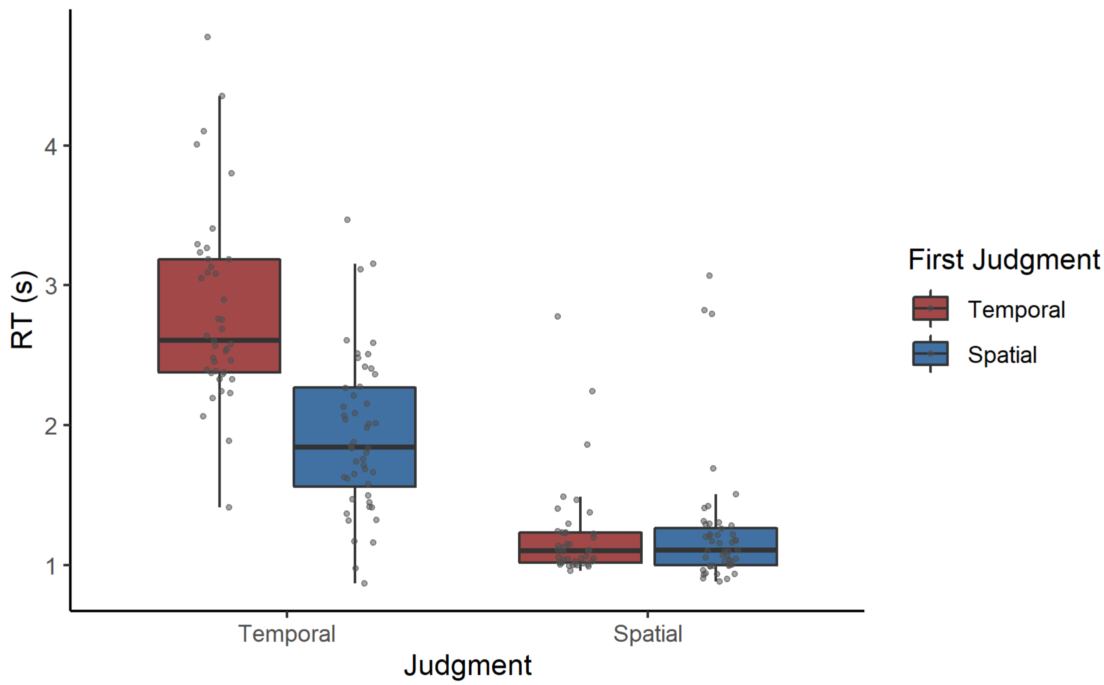

Figure 17. Behavioral Response Times in Experiment 4
Box plots indicate the median and quartile ranges of response times (RTs) associated with temporal and spatial judgments, dependent on whether the temporal or spatial judgment was made first (between-subjects).

5.2.2 Relating Spatial and Temporal Precision

Spatial errors were calculated as the Euclidean distance between the correct location and the responded location. As raw temporal judgments were calculated to include both positive and negative values, we first needed to transform the temporal errors to its absolute values to enable the comparison of errors across the two judgment tasks. Next, judgments (temporal and spatial) were standardized to a mean of zero and unit variance to ensure each judgment was on the same scale for testing. Figure 18 shows distributions of both tasks to visualize this transformation process. To provide a cursory examination of whether the scaled temporal and spatial judgments were correlated, RTs for the two were first averaged for each subject. Figure 19 shows a scatterplot between spatial and temporal precision, along with the estimated line of best fit. There was a significant positive correlation between the two types of judgments, r = .38, $t(109)$ = 4.27, p < .001. As it was important to confirm this statistical relationship at the trial level, we confirmed this relationship was significant when correlating spatial and temporal RTs in an unaveraged multilevel modeling framework, $t(7839.86)$ = 2.92, p = .003.

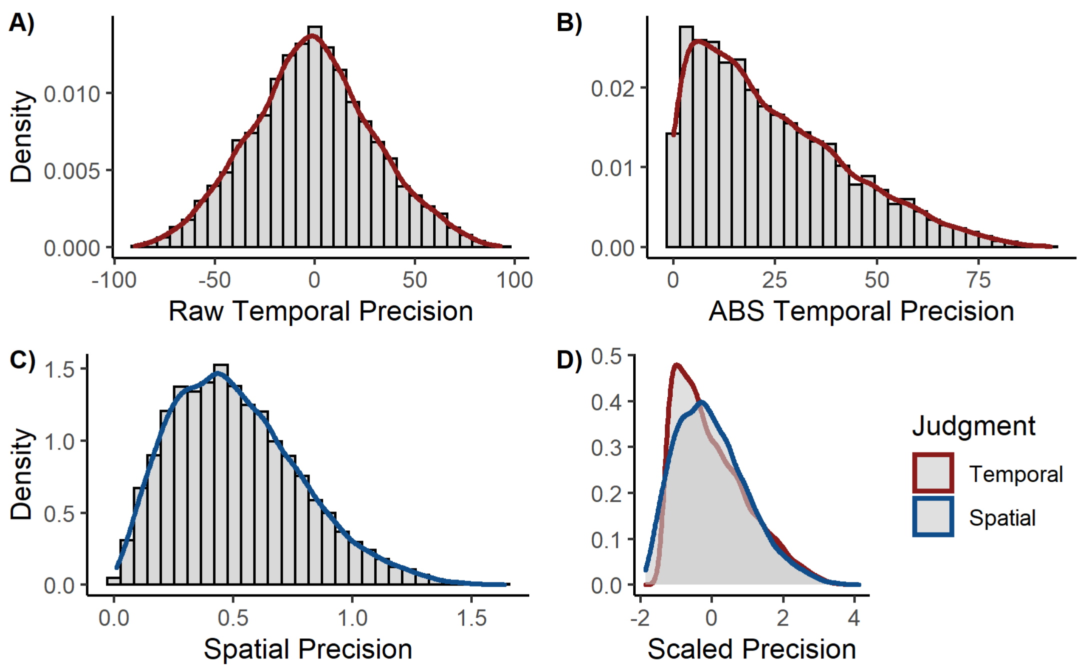

Figure 18. Error Distributions of Temporal and Spatial Judgments in Experiment 4

(A) Density plot displaying the distributions of error in temporal judgments, centered on each item's correct position.
(B) Distribution of absolute-valued temporal judgment errors, which were computed for better comparison with the spatial judgments.
(C) Density plot displaying the distributions of error (Euclidean distance) in spatial judgments, centered on each item's correct position.
(D) Temporal and spatial error distributions are superimposed after standardizing each judgment to a mean of zero with unit variance.

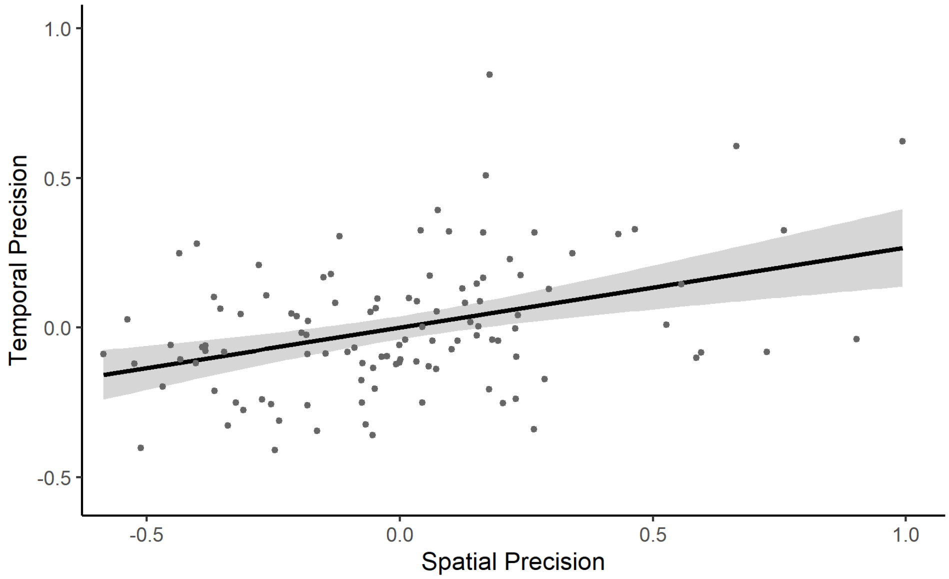

Figure 19. Regression Results of Temporal and Spatial Precision in Experiment 4
The fitted regression line indicates the relationship between temporal and spatial precision. Each dot corresponds to data for an individual subject.

Next, we tested for any differences in the errors associated with judging temporal and spatial memory. To compare errors across the two tasks, the absolute values of temporal errors were calculated, and then each judgment (temporal and spatial) was scaled to a mean of zero and unit variance. In the process of comparing temporal and spatial memory, a factor corresponding to the between-subjects manipulation of judgment order (temporal first or spatial first) was also included in the analysis. (We had no predictions regarding the latter factor, as it merely served for counterbalancing.) Figure 20 shows the medians of the scaled error distributions for each condition. The resulting 2 × 2 mixed ANOVA revealed no main effect of spatial versus temporal errors, $F(1, 15843.3) = 0.04$, $p = .839$, as well as no main effect of judgment order (e.g., spatial versus temporal first), $F(1, 109.3) = 3.02$, $p = .085$. There was, however, a significant interaction between judgment type and order, $F(1, 15843.3) = 7.99$, $p = .005$. Whereas temporal error did not change depending on whether subjects made this judgment first ($M = 0.02$, $SD = 1.02$) or second ($M = -0.01$, $SD = 0.98$), $t(110.04) = 0.69$, $p = .493$, spatial error was lower when made first ($M = -0.05$, $SD = 0.97$) versus second ($M = 0.06$, $SD = 1.03$), $t(109.08) = 2.00$, $p = .048$, suggesting that subjects had difficulties precisely retrieving the spatial location of an object after making a retrieval judgment in a different domain.

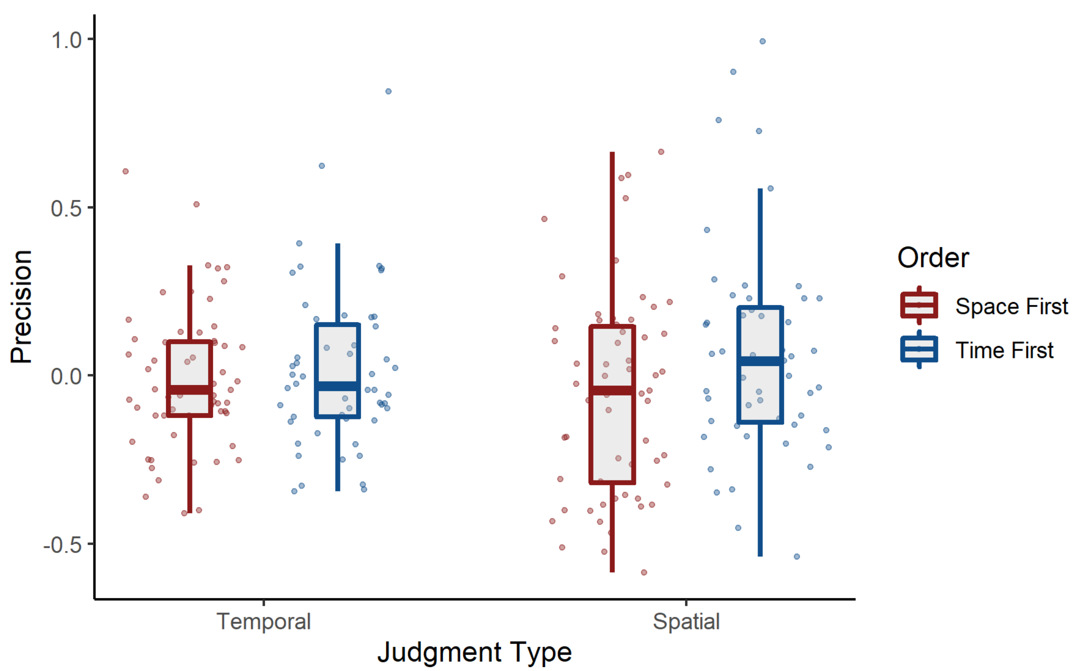

Figure 20. Transformed Temporal and Spatial Precision Results in Experiment 4
Box plots show the median and quartile range of precision values for temporal and spatial judgments, which have been transformed (z-scaled) for comparison. The results are also separated according to whether subjects completed the temporal or spatial judgment first.

Considering this interaction difference, we further assessed whether the second judgment made by a subject was correlated with the first judgment, dependent on the order in which subjects completed them. In subjects that made the spatial judgment first, spatial error was positively correlated, and significantly so, with the error for the subsequent temporal judgment, $b = 0.04$, $t(57.66) = 2.58$, $p = .013$, which could indicate that precisely retrieving the spatial location of an object facilitated its temporal retrieval, or involved an ease of switching between these two modes of episodic retrieval. Alternatively, conditions where temporal judgments were provided first, temporal errors did not significantly predict the immediately following spatial judgment, $b = 0.02$, $t(48.24) = 0.90$, $p = .371$. Figure 21 shows scatter plots between temporal and spatial precision, split on which judgment was provided first. Figure 21, along with the group level regression coefficients revealed that both conditions showed a positive relationship. However, the inference tests, along with Figure 22 showing regression coefficients at the subject level suggests that this relationship was not evident for when temporal judgments were provided first. This could suggest that spatial judgments are not influenced by retrieving other aspects of an item, or that subjects experienced more difficulty in switching into the mode required to precisely retrieve spatial information.

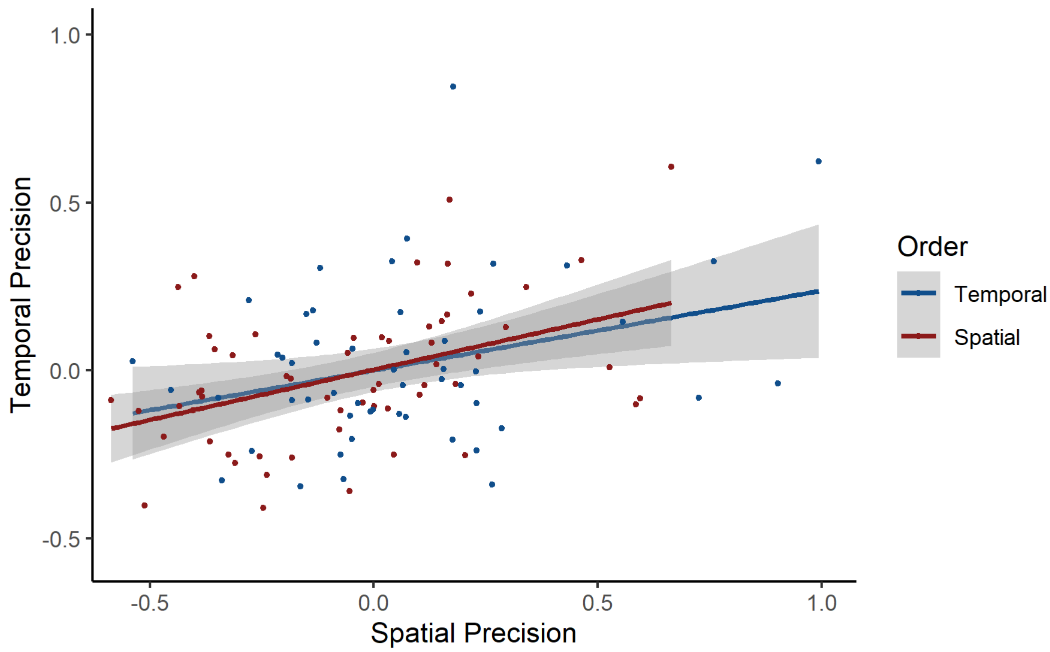

Figure 21. Regression Results of Temporal and Spatial Precision by Judgment Order in Experiment 4
The fitted regression lines indicating the relationship between temporal and spatial precision, separate according to which judgment subjects made first. Each dot corresponds to data for an individual subject.

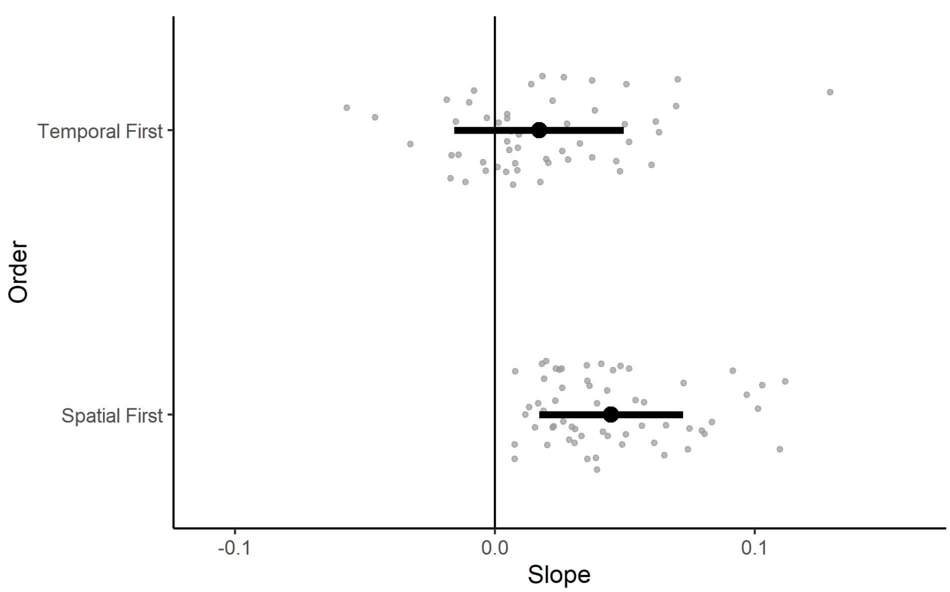

Figure 22. Subject-Wise Regression Results of Temporal and Spatial Precision by Judgment Order in Experiment 4

Forest plot displaying regression coefficients of temporal and spatial precision, according to which judgment was first. Black circles represent the group-level estimate, the error bars correspond to the standard deviation, and the subject-wise estimates are in grey.

5.2.3 Recency Effects on Temporal and Spatial Precision

In line with the first three experiments, we finally assessed how precision for both temporal and spatial judgments changed according to how long ago an item was studied. Recall that in the first two of three experiments, we observed strong evidence of a recency effect in temporal precision. Here, recency effects were statistically tested through trial level regressions using encoding position as a predictor. The results of these analyses are shown in Figure 23. Similar to the first two experiments, we replicated the temporal recency effects using a comparable length of encoding list. Items studied further in the past were associated with worse temporal precision, $t(7879) = 2.30$, $p = .022$. The critical test was to novelly assess if, like temporal precision, spatial precision showed the same type of degradation over time. Strikingly, this was not the case for spatial precision. Across the encoding period, spatial precision remained at similar levels, $t(7872) = 0.34$, $p = .733$, and did not show the typical recency patterns found in temporal precision. As observed in prior experiments, subjects are retrieving temporal information with higher precision for items that were more recently studied, but this information is vulnerable to degradation with passing time. Spatial precision appears to remain more stable and consistent across the study period, and may not be vulnerable to the same types of disruption that influences precise retrieval.

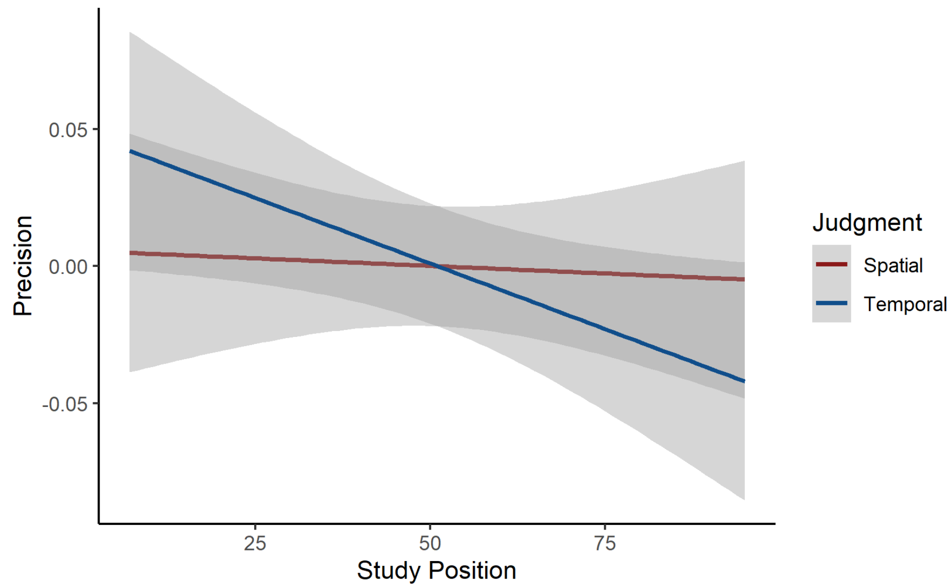

Figure 23. Recency Effects of Temporal and Spatial Precision in Experiment 4

The group-level regression lines are plotted for temporal (blue) and spatial (red) precision as a function of recency (study position).

6. GENERAL DISCUSSION

The overall aim of the experiments reported here was to further understand the precision with which temporal information about past experiences is retrieved. Retrieval precision, whether it is of interest in the context of short-term/working memory or long-term memory, is commonly investigated with the continuous report task. A common version of this task, as described in the General Introduction, involves having subjects associate a series of items during encoding with spatial positions designated along a circle. Then, on a later memory test, the items are used as cues to recall and judge, as precisely as possible, the corresponding location. Importantly, the retrieval task provides a continuous distribution of errors between the correct and judged locations, allowing for statistical mixture-modeling the precision of retrieval, along with the involvement of other decision-related (i.e. non-mnemonic) factors. Across four experiments, we adapted this approach of assessing retrieval precision to the domain of temporal information, which is a ready target for investigation given its prevalent association with episodic memory and its ever-changing nature. The general task employed here involved having subjects incidentally encode a series of items and then subsequently estimate the temporal position of each item. By novelly applying mixture-modeling to continuous temporal judgments, and expanding this approach to multilevel (hierarchical) regression, the findings of recent studies on temporal memory were extended to not only characterize precision and how it changes with passing time, but to also test for the influence of decision-making processes such as guessing and bias. Below, we summarize and interpret the main results that are specific to each experiment, while drawing more general theoretical conclusions from the consistent findings across experiments when possible.

Experiment 1 introduced the paradigm for continuously judging encoding time and provided motivation for the utility of investigating the precision of retrieving temporal information in service of episodic memory. Most relevant to our main hypothesis was the finding that temporal precision, parameterized as the standard deviation of a normal function fit to the error distribution, was better for items encoded more recently. This recency effect is especially notable given its reliability within the bounds of the encoding list, which lasted only 17.5 minutes. The effect was apparent both from our main analysis that divided the list in half—giving rise to an average remote-recent time difference of about 6 minutes, based on the midpoint of each half—and from a secondary, quartile-based analysis. While the latter analysis is less optimal in terms of trial numbers per bin, it afforded the ability to distinguish differences in precision between the middle and ends of the list from the recency effect, which was evident even between the middle quartiles (~3 minutes between midpoints). Together, this evidence extends that of previous studies in which estimates of retrieval precision, for either continuous color (Brady et al., 2013) or spatial location (Harlow & Donaldson, 2013; Sutterer et al., 2019), have also been shown to decline, but with considerably longer delays between encoding and retrieval. The convergence across these domains also informs a much more expansive literature on forgetting (Singh, Oliva, & Howard, 2017; for reviews, see Friedman, 1993; Wixted, 2004; Howard et al., 2015), further specifying the decline in accuracy with remoteness as diminished precision rather than increased guessing.

The involvement of random guessing, a non-mnemonic factor, in temporal judgments was tested in Experiment 1 by including a uniform distribution in the mixture-modeling procedure. We observed minimal influence of guessing of this sort, contrary to that found in numerous prior studies (e.g., Zhang & Luck, 2008; Brady et al., 2013; Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016), which could be explained in multiple but likely related ways. On one hand, whereas guessing along circular representations of location and color has been successfully modeled this way, our subjects tended to respond less frequently near the ends of the list, similar to previous observations using a timeline (Jenkins & Ranganath, 2010) and near the edges of two-dimensional space (e.g., Nilakantan et al., 2018). Thus, there might be something unique about responding along a linear scale, as is the case with time. On the other hand, an alternative explanation for the discrepancy is that, beyond any property of our response method, memory for time is inherently distinct from that for other features. That is, for any given memory, there could invariably be some temporal information that can be retrieved, albeit vague; for instance, it might always be possible to place a memory in the correct as opposed to incorrect half of an encoding list, even when precision is rather low. The nature of cognitive processes underlying these vague (i.e. acontextual) memories, as well as whether guessing is better captured by other (nonuniform) functions, are critical issues to address in future studies.

The final set of findings from Experiment 1 worthy of discussion are related to response bias. First, there was an overall tendency to judge encoded items as occurring more recently than they actually did. This finding is consistent with some studies of judgments of recency (e.g., Hintzman & Block, 1971; McCormack, 1984), including a subset using absolute judgments that are analogous to our task (Hinrichs & Buschke, 1968; Lockhart, 1969; Linton, 1975; Underwood, 1977). Notably, whereas those studies primarily tested memory over either short or very long periods (i.e. seconds versus months/years), our design aligns better with the vast majority of modern studies of long-term memory, where encoding-retrieval delay is typically on the order of tens of minutes. In the context of modeling retrieval precision in long-term memory, the current study is to our knowledge novel in testing for response bias. A design such as this could thus provide an important bridge for integrating temporal judgments, as well as their associated biases, into existing episodic memory theories. The second finding was that temporal judgments made with high confidence were associated with an elevated level of recency bias compared to those made with low confidence. Admittedly, there were no *a priori* predictions about how subjects might arrive at judging confidence. Whereas the instructions emphasized that it should be treated separately from precision (Harlow & Yonelinas, 2016), it seemed possible that subjects would resort to relying on precision in the absence of any other evidence. Alternatively, the basis for confidence could have varied from subject to subject, resulting in the measure being just a noisier version of the precision judgment (also see Harlow & Donaldson, 2013). Nonetheless, it is plausible that the bias effect we observed is due to some subjective factor, such as familiarity strength, that gives rise to increases in both confidence and a feeling of recency.

Experiment 2 served primarily as a conceptual replication of the main, and for the most part novel, findings from the first experiment. The experimental paradigm was largely the same, with one notable change corresponding to the encoding period being roughly halved (140 items compared to the 300 used previously). Regarding the findings from the mixture-modeling analyses that were analogous to those of Experiment 1, the effect of response bias and the negligible role of guessing were both replicated. Moreover, the analysis approach of the previous experiment was extended by employing a hierarchical regression-based version of the mixture-modeling procedure. This allowed us to capitalize on statistical power afforded by assessing continuous changes across the encoding phase, as opposed to relying on the binning of trials into discrete groups. The latter analysis indicated that precision decreased linearly on a trial-by-trial basis with passing time, providing converging support for the recency effects observed in Experiment 1.

An additional purpose of Experiment 2 was to test for differences in temporal retrieval precision according to whether the tested pictures came from animate versus inanimate categories. As has been demonstrated in several recent studies employing a range of tasks, there is a growing body of evidence that animate stimuli (e.g., a bear, bird, or child) are successfully retrieved at a higher rate than inanimate stimuli (e.g., a chair, hammer, or bowl), consistent with an evolutionary view in which human memory is functionally adaptive (Laurino & Kaczer, 2019; Nairne et al., 2013; Nairne et al., 2017; VanArsdall et al., 2013; VanArsdall et al., 2015). However, when testing for such categorical differences with the current paradigm, we observed an invariance with respect to the precision of temporal judgments. While it remains to be determined whether our experimental design represents a boundary condition for adaptive memory, the findings are nonetheless informative in terms of replicating the results of Experiment 1 across two stimulus categories. Thus, while researchers often aim to control stimulus factors such as imageability, frequency, and meaningfulness when choosing experimental stimuli, animacy does not appear to be of concern when considering temporal information.

For Experiment 3, the main differences in methodology from the first two experiments was that a much longer encoding list was employed and EEG data were acquired to identify neural correlates of temporal retrieval precision. In particular, this experiment increased the number of encoding trials five-fold relative to Experiment 1 and over ten-fold compared to Experiment 2. One finding that replicated with the longer encoding list was the negligible role of guessing in response errors. In particular, the results from model comparison pointed to a better fit when excluding the uniform guessing component. As laid out in the Introduction to Experiment 3, this additional test of guessing was important, as the longer length of the encoding phase allowed for analyzing enough trials near the middle of the list that were not affected by the reduced response rates at the ends. Thus, consistent with the findings of the previous experiments, guessing appears to minimally influence judgments of temporal retrieval.

One other notable aspect of Experiment 3’s findings was that the recency effect on precision was not observed when binned trials were modeled separately, contrasting with the findings from the first two experiments. Further, the hierarchical trial-level analyses revealed an effect in the opposite direction as expected, such that more recent items were associated with *worse* precision than remote items. While the main difference between experiments was the length of the encoding period, one explanation of the discrepant findings is that there was an effect of fatigue in the current experiment, where any benefit to recency was washed out by the substantially longer list. Some evidence for this idea comes from the fact that precision was much worse in the current than previous experiments. Alternatively, it is also possible that the switch to more memorable picture stimuli, as opposed to simple object pictures, had an effect on recency. In particular, the pictures used in Experiment 3 might have resulted in an increase in the saliency of their inherent details, thus enhancing item memorability at the expense of memory for temporal context.

Turning to the electrophysiological data from Experiment 3, the analyses initially focused on the late parietal effect (LPE), a positive ERP component that typically appears around 500-800 ms post-stimulus onset over left-lateralized parietal electrodes, and which has been studied extensively in the context of episodic memory retrieval (Smith, 1993; Friedman & Johnson, 2000; Rugg & Curran, 2007). While splitting test trials into two groups corresponding to high and low precision resulted in no difference in the LPE, binning into quartiles according to precision revealed a marginal effect in which ERP amplitude was positively correlated with precision. This finding is consistent with that of a previous study by Murray et al. (2015) in which the LPE amplitude changed with retrieval precision on a spatial (circular) task. Building on this result, we next tested for precision-related modulation of ERP effects in a more continuous and exploratory manner using a multilevel regression approach that was applied separately to each electrode and 25-ms time bin across the recording epoch. Again, we observed a significant effect of precision around the 500-800 ms interval, but this effect was more prominent over right central as opposed to left parietal scalp. Although we extend the evidence for ERP correlates of retrieval precision to temporal information, since this experiment constitutes the first such test of this effect, it remains to be determined whether it replicates in a larger sample of subjects and if its topographic distribution is indeed distinct from that for spatial memory.

In two additional sets of analyses of the EEG data, frequency power in the alpha and theta bands were examined in relation to the behavioral measures of temporal precision. Regression-based analyses revealed a late-onsetting (1100-1400 ms) effect in the alpha band over left parietal sites, whereas there were two effects in the theta band: one over left central scalp at 900-1100 ms and the other over left frontal scalp at 1600-1700 ms. Since these results come from exploratory analyses, it is important to be cautious in making any inferences about their functional significance. Perhaps one interesting aspect of the results, however, is that they largely occur beyond the time frame of typical effects of episodic retrieval such as the 500-800 ms interval of the LPE. Instead, the timing of these oscillatory effects aligns better with EEG correlates of post-retrieval monitoring (Koriat & Goldsmith, 1996; Wilding & Rugg, 1996; Hayama et al., 2008), suggesting that they might reflect the maintenance or further evaluation of the temporal information retrieved.

In the final experiment reported here, we sought to place our findings for precisely retrieving temporal information in the larger context of analogous studies that more commonly test spatial retrieval. After encoding a series of pictures that were displayed at random locations on the computer screen, subjects made both a temporal and spatial judgment for each stimulus at test. In one series of analyses, we sought to investigate the correspondence between these two types of judgment. Overall, the judgments were positively correlated at the trial level, such that trials on which temporal accuracy was high also exhibited accurate spatial retrieval. In addition, when the distributions of temporal and spatial errors were *z*-adjusted for comparison, there was no observable difference in precision between the two. Finally, although we did not expect prior to conducting Experiment 4 that counterbalancing the order of temporal and spatial judgments would have an effect on dependencies, the temporal-spatial relationship was stronger for subjects who made spatial judgments first. Nevertheless, when considered together, these findings provide initial and somewhat consistent evidence that temporal and spatial retrieval might share some underlying neurocognitive processes.

Experiment 4 also provided a further test of the recency effects on temporal precision that were evident in Experiments 1 and 2 (cf. Experiment 3). Using a shorter encoding list compared to those experiments (100 versus 300 and 140 items), we again replicated the result that temporal judgments were more precise for recent compared to remote trials. However, while precision was largely correlated across domains, as described above, spatial judgments did not exhibit the recency effect that temporal judgments did. One possible explanation of this dissociation is that the retrieval of temporal information is uniquely supported by a vague sense of relative item strength, whereby the strength of recent items declines as those items become more remote. The same sort of strength-based signal is uninformative, though, to judgments of spatial retrieval, as that factor varied randomly across the changing temporal context of the encoding phase. As this study provides only a starting data point in contrasting the dynamics of spatial and temporal memory, further investigations are clearly necessary to test the explanation put forth here, as well as alternative ideas about whether forgetting differs across domains.

In conclusion, just as recent studies investigating memory precision have advanced working memory theories beyond a predominant focus on capacity (for reviews, see Cowan, 2001; Luck & Vogel, 2013), the continuous report paradigm is also proving to be instrumental in moving episodic memory research past the “all-or-none”, thresholded form of recollection (Parks & Yonelinas, 2009; Wixted, 2007). In the set of experiments reported here, the employment of this paradigm, along with the corresponding mixture-modeling approach to analysis, was intended to capitalize on subtle changes in the precision of memory retrieval. Characterizing these changes with statistically-sensitive techniques is the first step to understanding how continuous retrieval judgments are the result of neural processes that might also operate in a graded manner (e.g., Johnson et al., 2009; Murray et al., 2015). Understanding and characterizing the way in which we retrieve temporal information has several potential applications. As we elucidate these processes in healthy young adults, it will likewise be informative to begin to understand these processes in healthy ageing, as well as if temporal memory precision is affected in neurodegenerative disorders such as Alzheimer's, and to the extent temporal retrieval is damaged compared to other types of memory (e.g., basic recognition, spatial memory). For example, if signs of temporal memory degradation occurs earlier than say, spatial memory or basic recognition, this could be informative in early detection of aforementioned neurodegenerative disorders. Similarly, future research could aim at assessing the importance and relationship between temporal memory precision and educational achievement. Memory performance has been correlated with other cognitive factors that contribute to educational achievement. It would be fruitful to investigate how these types of memory paradigms can influence and direct aspects of educational programs, and ultimately, academic achievement. Generally speaking, as we have demonstrated here, time has the potential to be a useful dimension for investigating such phenomena, given its ever-present association with encoded events, even in the absence of any overt experimental manipulations, and its changing dynamics even within single-session laboratory tests of episodic memory.

REFERENCES

Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools*, *49*(3), 340–355. doi:10.1044/2018_lshss-17-0114

Andrieu, C., De Freitas, N., Doucet, A., & Jordan, M. (2003). An introduction to MCMC for machine learning. *Machine Learning*, *50*(1), 5–43. doi:10.1023/A:1020281327116

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. doi:10.1167/9.10.7

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal Feature Integration in the Angular Gyrus during Episodic and Semantic Retrieval. *The Journal of Neuroscience*, *36*(20), 5462–5471. doi:10.1523/JNEUROSCI.4310-15.2016

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981–990. doi:10.1177/0956797612465439

Burkner, P. C. (2017). Brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1-28. doi:10.18637/jss.v080.i01

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. doi:10.1017/S0140525X01003922

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*, 657-661. doi:10.1038/nature08704

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, *95*(5), 1007-1018. doi:10.1016/j.neuron.2017.06.036

Ekstrom, A. D., & Yonelinas, A. P. (2020). Precision, binding, and the hippocampus: Precisely what are we talking about? *Neuropsychologia*, *138*(17), 107341. doi:10.1016/j.neuropsychologia.2020.107341

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*(1). doi:10.1038/ncomms2237

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, *115*(1), 168-177. doi:10.1152/jn.00860.2015

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*(1), 44-66. doi:10.1037/0033-2909.113.1.44

Friedman, D., & Johnson, R. J. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research & Technique*, *51*(1), 6-28. doi:10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4), 457–472. doi:10.1214/ss%2F1177011136

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., … Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791-807.e7. doi:10.1016/j.neuron.2017.07.011

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446-460. doi:10.1016/j.neuroimage.2013.10.027

Gronau, Q. F., & Singmann, H. (2018). bridgesampling: Bridge sampling for marginal likelihoods and bayes factors. R package version 0.6-0. https://CRAN.R-project.org/package=bridgesampling

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, *20*(2), 318–325. doi:10.3758/s13423-012-0340-9

Harlow, I. M., & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory* , *24*(1), 114–127. doi:10.1080/09658211.2014.988162

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: Specific or non-specific? *Neuropsychologia*, *46*(5), 1211-1223. doi:10.1016/j.neuropsychologia.2007.11.021

Hayes, S. M., Buchler, N., Stokes, J., Kragel, J., & Cabeza, R. (2011). Neural correlates of confidence during item recognition and source memory retrieval: Evidence for both dual-process and strength memory theories. *Journal of Cognitive Neuroscience*, *23*(12), 3959-3971. doi:10.1162/jocn_a_00086

Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Sciences*, *24*(3), 208-227. doi:10.1016/j.tics.2019.12.006

Hinrichs, J. V., & Buschke, H. (1968). Judgment of recency under steady-state conditions. *Journal of Experimental Psychology*, *78*(4, Pt.1), 574-579. doi:10.1037/h0026615

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297-306. doi:10.1037/h0030907

Howard, M. W. (2017). Temporal and spatial context in the mind and brain. *Current Opinion in Behavioral Sciences*, *17*, 14-19. doi:10.1016/j.cobeha.2017.05.022

Howard, M. W., Shankar, K. H., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24-53. doi:10.1037/a0037840

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience*, *30*(46), 15558-15565. doi:10.1523/JNEUROSCI.1337-10.2010

Johnson, M. K., Hashtroudi, S., & Lindsay, S. D. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67-85. doi:10.1037/0033-295X.88.1.67

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex*, *17*(11), 2507-2515. doi:10.1093/cercor/bhl156

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, *63*(5), 697-708. doi:10.1016/j.neuron.2009.08.011

Jung, T. -P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745-1758. doi:10.1016/s1388-2457(00)00386-2

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. *International Conference on Computer Vision (ICCV)*. doi:10.1109/ICCV.2015.275

Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, *50*(4), 1648–1657. doi:10.1016/j.neuroimage.2010.01.051

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., … Moser, M. -B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, *321*(5885), 140-143. doi:10.1126/science.1157086

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490-517. doi:10.1037/0033-295X.103.3.490

Kraus, B. J., Robinson, R. J., 2nd, White, J. A., Eichenbaum, H., Hasselmo, M. E. (2013). Hippocampal "time cells": Time versus path integration. *Neuron*, *78*(6), 1090-1101. doi:10.1016/j.neuron.2013.04.015

Laurino, J., & Kaczer, L. (2019). Animacy as a memory enhancer during novel word learning: Evidence from orthographic and semantic memory tasks. *Memory*, *27*(6), 820-828. doi:10.1080/09658211.2019.1572195

Linton, M (1975). Memory for real-world events. In D. A. Norman and D. E. Rumelhart (Eds.), *Explorations in Cognition*. W. H. Freeman and Company, San Francisco: CA

Lockhart, R. S. (1969). Recency discrimination predicted from absolute lag judgments. *Perception & Psychophysics*, *6*, 42-44. doi:10.3758/BF03210673

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *eLife*, e16070, doi:10.7554/eLife.16070

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391-400. doi:10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*, 347-356. doi:10.1038/nn.3655

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271. doi:10.1037/0033-295X.87.3.252

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419-457. doi:10.1037/0033-295x.102.3.419

McCormack, P. D. (1984). Aging and recognition memory: Methodological and interpretive problems. *Experimental Aging Research*, *10*(4), 215-219. doi:10.1080/03610738408258468

Mickes, L., Wais, P., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science*, *20*(4), 509-515. doi:10.1111/j.1467-9280.2009.02324.x

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What we have learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638-677. doi:10.1037/a0015849

Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*(5), 504-520. doi:10.1002/hipo.20178

Montchal, M. E., Reagh, Z. M., & Yassa, M. A. (2019). Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nature Neuroscience*, *22*, 284-288. doi:10.1038/s41593-018-0303-1

Morey, R. D., Rouder, J. N., & Jamil, T. (2018). BayesFactor: Computation of Bayes Factors for common designs. *R package version* 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Murray, J. G., Howie, C. A., & Donaldson, D. I. (2015). The neural mechanism underlying recollection is sensitive to the quality of episodic memory: Event related potentials reveal a some-or-none threshold. *NeuroImage*, *120*, 298–308. doi:10.1016/j.neuroimage.2015.06.069

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*(1), 22-27. doi:10.1177/0963721416667711

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099-2105. doi:10.1177/0956797613480803

Nilakantan, A. S., Bridge, D. J., Gagnon, E. P., VanHaerents, S. A., & Voss, J. L. (2017). Stimulation of the Posterior Cortical-Hippocampal Network Enhances Precision of Memory Recollection. *Current Biology*, *27*(3), 465–470. doi:10.1016/j.cub.2016.12.042

Nilakantan, A. S., Bridge, D. J., VanHaerents, S., & Voss, J. L. (2018). Distinguishing the precision of spatial recollection from its success: Evidence from healthy aging and unilateral mesial temporal lobe resection. *Neuropsychologia*, *119*, 101–106. doi:10.1016/j.neuropsychologia.2018.07.035

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1023-1035. doi:10.1016/j.neubiorev.2009.12.014

Parks, C. M., & Yonelinas, A. P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11515-11519. doi:10.1073/pnas.0905505106

Peirce, J. W. (2007). PsychoPy - Psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13. doi:10.1016/j.neumeth.2006.11.017

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230–240. doi:10.1016/j.tics.2013.03.005

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rajaram, S. Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89-102. doi:10.3758/BF03211168

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews. Neuroscience*, *13*(10), 713–726. doi:10.1038/nrn3338

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife*, *5*. doi:10.7554/eLife.18260

Ritchey, M., Libby, L. A., & Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. *Progress in Brain Research*, *219*, 45–64. doi:10.1016/bs.pbr.2015.04.001

Rouder, J. N., Engelhardt, C. R., McCabe, S., & Morey, R. D. (2016). Model comparison in ANOVA. *Psychonomic Bulletin & Review*, *23*(6), 1779-1786. doi:10.3758/s13423-016-1026-5

Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 19-26. doi:10.1177/2515245917745058

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis.

Psychonomic Bulletin & Review, *16*(2), 225–237. doi:10.3758/PBR.16.2.225.

Rugg, M. D., Cox, C. J. C., Doyle, M. C., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia*, *33*(4), 471-484. doi:10.1016/0028-3932(94)00132-9

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. doi:10.1016/j.tics.2007.04.004

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260. doi:10.1016/j.conb.2012.11.005

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia*, *50*(13), 3070–3079. doi:10.1016/j.neuropsychologia.2012.06.004

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., & Eichenbaum, H. (2016). Time cells in hippocampal area CA3. *Journal of Neuroscience*, *36*(28), 7476-7484. doi:10.1523/jneurosci.0087-16.2016

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657-661. doi:10.1038/nrn2213

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464.

Schwarzkopf, D. S. (2015). Bayes Factors MATLAB functions. Figshare. https://figshare.com/articles/Bayes_Factors_Matlab_functions/1357917/1. doi:10.6084/m9.figshare.1357917.v1

Scofield, J. E., Price, M. H., Flores, A., Merkle, E. C., & Johnson, J. D. (2020). Repetition attenuates the influence of recency on recognition memory: Behavioral and electrophysiological evidence. *Psychophysiology*, *57*(9), e13601. doi:10.1111/psyp.13601

Singh, I., Oliva, A., & Howard, M. (2017). Visual memories are stored along a compressed timeline. bioRxiv. https://www.doi.org/10.1101/101295

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging patient data. *Neuropsychologia*, *45*(10), 2163-2179. doi:10.1016/j.neuropsychologia.2007.03.007

Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience*, *5*(1), 1-13. doi:10.1162/jocn.1993.5.1.1

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-*N* design. *Psychonomic Bulletin & Review*, *25*, 2083-2101. doi:10.3758/s13423-018-1451-8

Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia*, *47*(8-9), 1765–1779. doi:10.1016/j.neuropsychologia.2009.02.028

Staresina, B. P., & Wimber, M. (2019). A neural chronometry of memory recall. *Trends in Cognitive Sciences*, *23*(12), 1071-1085. doi:10.1016/j.tics.2019.09.011

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10):9, 1–8. doi:10.1167/13.10.9

Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K., & Awh, E. (2019). Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of Neurophysiology*, *122*(2), 539-551. doi:10.1152/jn.00268.2019

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*, 57-62. doi:10.1038/s41586-018-0459-6

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

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*(4), 385-398. doi:10.1037/0003-066X.40.4.385

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1-25. doi:10.1146/annurev.psych.53.100901.135114

Underwood, B. J. (1977). *Temporal codes for memories: Issues and problems.* Lawrence Erlbaum.

Valentine, K. D., Buchanan, E. M., Scofield, J. E., & Beauchamp, M. T. (2019). Beyond p values: Utilizing multiple methods to evaluate evidence. *Behaviormetrika*, *46*(1), 121-144. doi:10.1007/s41237-019-00078-4

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, *60*, 172-178. doi:10.1027/1618-3169/a000186

VanArsdall, J. E., Nairne. J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657-663. doi:10.1080/09658211.2014.916304

Vilberg, K. L., & Rugg, M. D. (2007). Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*, *45*(10), 2216-2225. doi:10.1016/j.neuropsychologia.2007.02.027

Waldhauser, G. T., Braun, V., & Hanslmayr, S. (2016). Episodic memory retrieval functionally relies on very rapid reactivation of sensory information. *The Journal of Neuroscience*, *36*(1), 251-260. doi:10.1523/JNEUROSCI.2101-15.2016

Wilding, E. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology*, *35*(1), 81-87. doi:10.1016/s0167-8760(99)00095-1

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, *119*(3), 889-905. doi:10.1093/brain/119.3.889

Wilken, P., & Ma, W.. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. doi:10.1167/4.12.11

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235-269. doi:10.1146/annurev.psych.55.090902.141555

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

Woroch, B., & Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Research*, *1317*, 180-191. doi:10.1016/j.brainres.2009.12.074

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441-517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002-3008. doi:10.1523/JNEUROSCI.5295-04.2005

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235. doi:10.1038/nature06860

Appendix

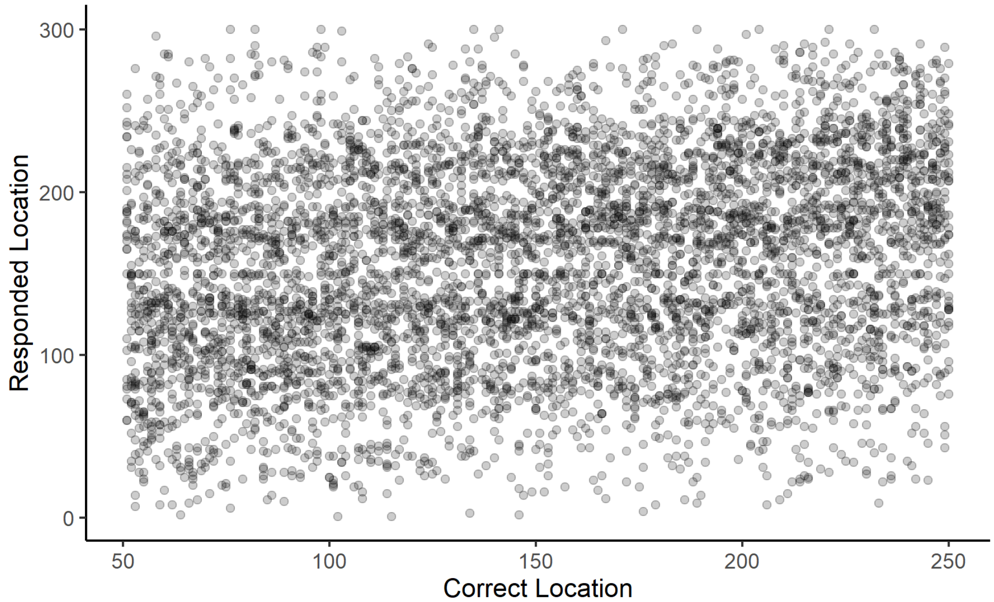

Figure A1. Scatterplot from Experiment 1 of the temporal judgments of each trial (y-axis) compared to its correct temporal study position (x-axis) across all subjects.

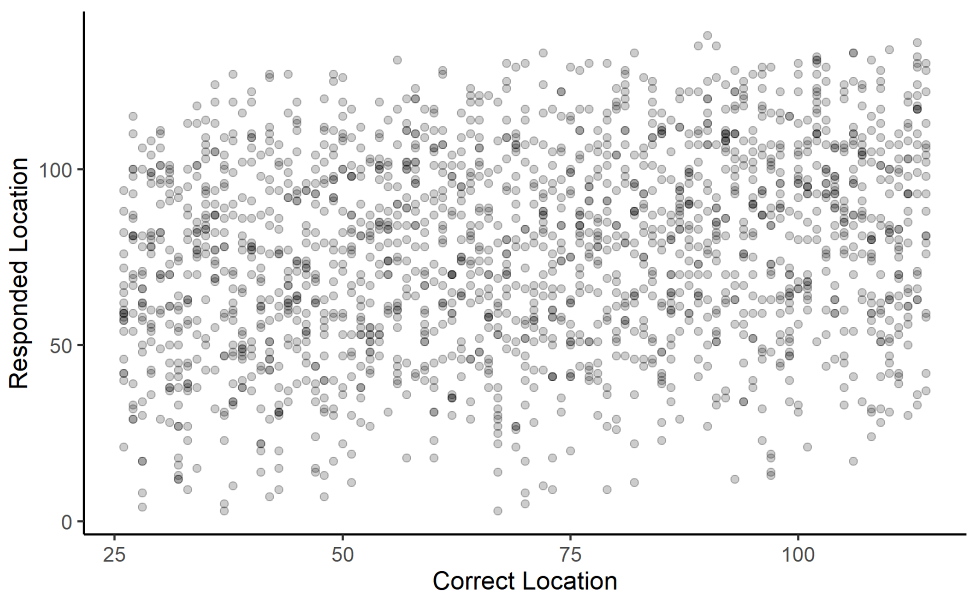

Figure A2. Scatterplot from Experiment 2 of the temporal judgments of each trial (y-axis) compared to its correct temporal study position (x-axis) across all subjects.

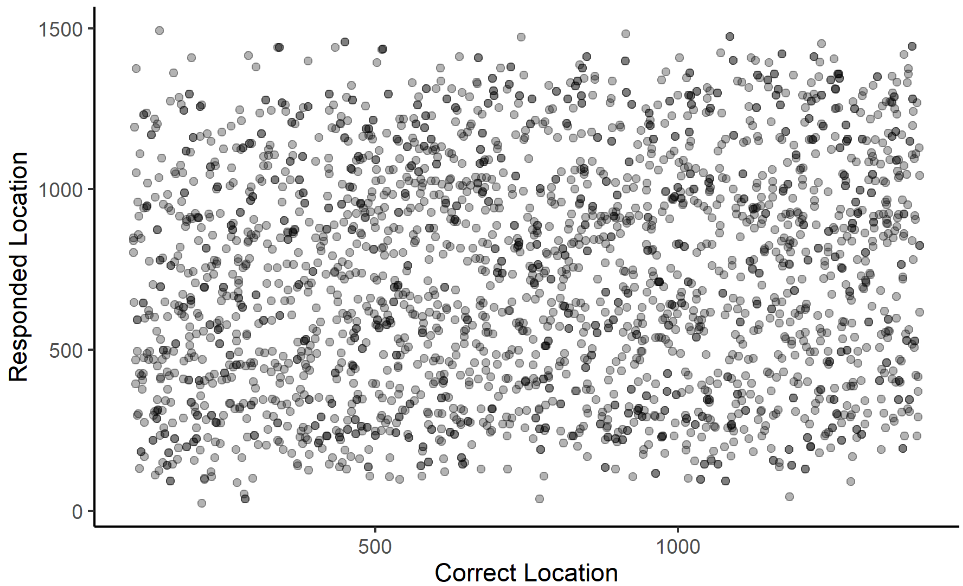

Figure A3. Scatterplot from Experiment 3 of the temporal judgments of each trial (y-axis) compared to its correct temporal study position (x-axis) across all subjects.

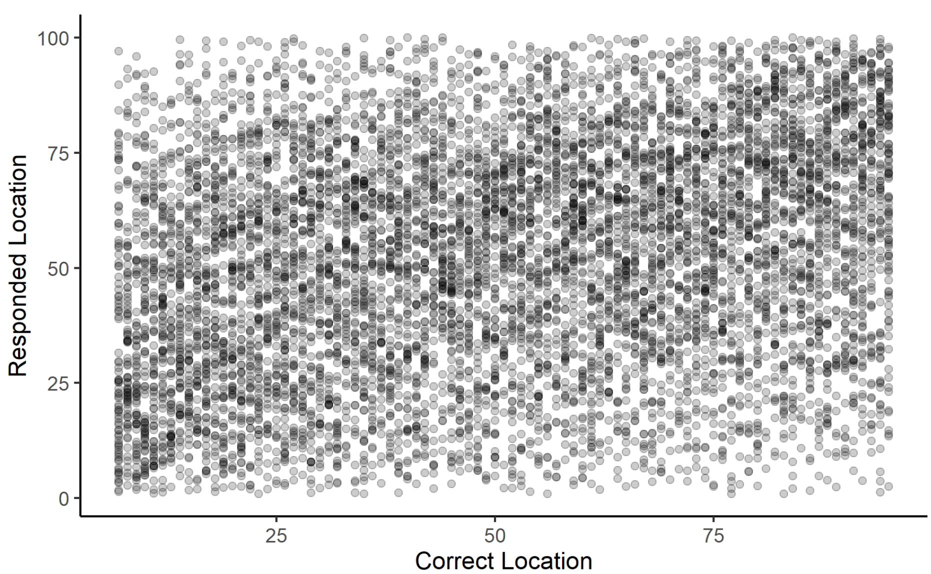

Figure A4. Scatterplot from Experiment 4 of the temporal judgments of each trial (y-axis) compared to its correct temporal study position (x-axis) across all subjects.

VITA

John E. Scofield was born in Temple, Texas in 1992. He graduated from Grandview High School in Grandview, Missouri before continuing his education at both Benedictine College in Atchison, Kansas, and then Truman State University in Kirksville, Missouri, where he studied and received a Bachelor of Science in Psychology. With a particular interest in Psychology, he next began graduate studies at Missouri State University in Springfield, Missouri. Studying under the guidance of Drs. Bogdan Kostic and Erin Buchanan, he mastered core concepts in Cognitive Psychology and Statistics, ultimately resulting in a Master of Science in Experimental Psychology, along with a Graduate Certificate in Statistics and Research Design. Further building on this core knowledge to extend to the field of cognitive neuroscience, he was selected as a PhD candidate in the Memory and Neuroimaging Laboratory at the University of Missouri at Columbia, studying under the guidance of Drs. Jeffrey Johnson and David Geary. There he further developed his skills in statistics and cognitive theory, and importantly gained critical insight on neuroimaging technologies and methodology. The training he received at Missouri State University and the University of Missouri at Columbia allowed for a continual search for answers related to fundamental theories of cognitive neuroscience and applied data analysis.