

**Distributed practice and distributed representations:  
Investigating the spacing effect using EEG**

by

**M. V. Mollison**

B.A., Brandeis University, 2005

M.A., University of Colorado Boulder, 2010

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Psychology and Neuroscience  
2015

This thesis entitled:  
Distributed practice and distributed representations: Investigating the spacing effect using EEG  
written by M. V. Mollison  
has been approved for the Department of Psychology and Neuroscience

---

Prof. Tim Curran

---

Prof. Alice F. Healy

---

Prof. Albert Kim

---

Prof. Leaf Van Boven

---

Prof. Michael C. Mozer

Date \_\_\_\_\_

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #13-0462

Mollison, M. V. (Ph.D., Cognitive Neuroscience)

Distributed practice and distributed representations: Investigating the spacing effect using EEG

Thesis directed by Prof. Tim Curran

The spacing effect shows that studying information at distributed intervals leads to better long-term memory than continuously studying over the same cumulative amount of time. Prior research has not made explicit predictions about neural activity in the EEG domain that should occur at different study repetition lags under different spacing effect theories. One main purpose of this thesis is to make these predictions in terms of the neural activity expected under different spacing effect hypotheses, and to test them to the best of our ability. Knowledge about neural patterns underlying the spacing effect can shed light on why this effect occurs, which would support or challenge the hypotheses presented here as well as possibly provide a physiological grounding when considering the spacing effect in educational settings. (This is incomplete.)

## Acknowledgements

First, thanks to my advisor, Dr. Tim Curran, for being a remarkable mentor who was always available to discuss ideas about research and analyses, who inspired me to pursue interesting and novel research topics, and who supported me throughout my graduate training.

Thanks to many of the faculty associated with the Department of Psychology and Neuroscience at the University of Colorado Boulder, particularly those who I learned from and worked with over the years. Thanks as well to Dr. Michael Kahana, whose lab sparked my interest in research on human memory when I was an undergraduate.

Thanks to Dr. Tim Curran, Dr. Alice Healy, Dr. Albert Kim, Dr. Leaf Van Boven, and Dr. Michael Mozer for being on my dissertation committee.

I would like to acknowledge and thank Chris Bird and Krystin Corby, the professional research assistants who helped organize and run participants for my experiments. I also could have not completed research nearly so quickly if it were not for the undergraduate research assistants who have been part of the Curran Lab over the years.

Thanks to Dr. Jeremy Manning and Dr. Phillip Gilley for discussions about and assistance with EEG similarity analyses.

Thanks to my family and friends for their valuable support over the years. You will be pleased to know that I am excited and happy with where pursuing a graduate degree has led me. I would not have made it here without following that path and your encouragement along the way.

This work was supported by NSF grant SMA-1041755 to the Temporal Dynamics of Learning Center.

## Contents

### Chapter

<b>1</b>	<b>General introduction</b>	<b>1</b>
1.1	Overview of the spacing effect . . . . .	1
1.2	Background for evaluated hypotheses . . . . .	3
1.2.1	Deficient Processing . . . . .	3
1.2.2	Contextual Variability . . . . .	4
1.2.3	Study-Phase Retrieval . . . . .	5
1.2.4	Hypothesis interactions . . . . .	8
1.3	Formal models accounting for the spacing effect . . . . .	8
1.4	EEG studies of the spacing effect . . . . .	9
<b>2</b>	<b>Dependent measures</b>	<b>11</b>
2.1	Neural data . . . . .	12
2.2	Effects of attention and semantic processing . . . . .	12
2.3	Effects of memory retrieval and encoding . . . . .	14
2.4	Memory reinstatement and contextual variability assessed via neural similarity . . .	16
2.5	Summary . . . . .	17
<b>3</b>	<b>Experiment 1</b>	<b>19</b>
3.1	Method . . . . .	19
3.1.1	Participants . . . . .	19

3.1.2	Materials . . . . .	19
3.1.3	Design . . . . .	20
3.1.4	Procedure . . . . .	20
3.1.5	Electrophysiological recordings and data processing . . . . .	23
3.2	Results . . . . .	25
3.2.1	Behavioral results . . . . .	26
3.2.2	ERP results . . . . .	26
3.2.3	ERP discussion . . . . .	32
3.2.4	Time–frequency results . . . . .	33
3.2.5	Time–frequency discussion . . . . .	40
3.2.6	Similarity results . . . . .	42
3.2.7	Similarity discussion . . . . .	45
3.3	Experiment 1 discussion . . . . .	46
<b>4</b>	<b>Experiment 2</b>	<b>49</b>
4.1	Method . . . . .	50
4.1.1	Participants . . . . .	50
4.1.2	Materials . . . . .	50
4.1.3	Design . . . . .	50
4.1.4	Procedure . . . . .	50
4.1.5	Electrophysiological recordings and data processing . . . . .	52
4.2	Results . . . . .	52
4.2.1	Behavioral results . . . . .	52
4.2.2	ERP results . . . . .	53
4.2.3	ERP discussion . . . . .	58
4.2.4	Time–frequency results . . . . .	61
4.2.5	Time–frequency discussion . . . . .	66

4.2.6	Similarity results . . . . .	68
4.2.7	Similarity discussion . . . . .	70
4.3	Experiment 2 discussion . . . . .	71
<b>5</b>	<b>General discussion</b>	<b>74</b>
5.1	Summary of results . . . . .	74
5.1.1	ERP results . . . . .	74
5.1.2	Time–frequency results . . . . .	75
5.1.3	Similarity results . . . . .	76
5.2	Alignment with the hypotheses . . . . .	77
5.2.1	Deficient Processing . . . . .	77
5.2.2	Contextual Variability . . . . .	77
5.2.3	Study-Phase Retrieval . . . . .	77
5.3	Future directions . . . . .	78
<b>References</b>		<b>79</b>
<b>Bibliography</b>		<b>79</b>

## Figures

### Figure

3.1	Experiment 1: Exposure, study, and test phases . . . . .	21
3.2	The 128-channel HydroCel Geodesic Sensor Net <sup>TM</sup> used to measure the EEG. In this top-down schematic the participant's nose is toward the front. . . . .	24
3.3	N1 to words at electrode 58 (T5) and neighbors, analyzed window 122–222 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed <i>vs.</i> spaced Presentation 2 (P2) across time and electrodes (*); (d) analyzed means (error bars are SEM). The early negative peak is significantly larger for spaced compared to massed repetitions. . . . .	28
3.4	N400 to words at electrode Cz and neighbors, analyzed window 322–422 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed <i>vs.</i> spaced Presentation 2 (P2) across time and electrodes (*); (d) analyzed means (error bars are SEM). The negative peak at 400 ms is significantly smaller for massed compared to spaced repetitions. . . . .	30
3.5	LPC to words at electrode 77 and neighbors, analyzed window 496–696 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed <i>vs.</i> spaced Presentation 2 (P2) across time and electrodes (*); (d) analyzed means (error bars are SEM). The positive peak around 600 ms is significantly larger for massed compared to spaced repetitions. . . . .	31

3.6 Theta power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (*); (c) and (f) mean values for the two time windows (error bars are SEM). . . . .	34
3.7 Lower alpha power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (*); (c) and (f) mean values for the two time windows (error bars are SEM). . . . .	36
3.8 Upper alpha power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (*); (c) and (f) mean values for the three time windows (error bars are SEM). . . . .	37
3.9 Lower beta power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (*); (c) and (f) mean values for the three time windows (error bars are SEM). . . . .	39
3.10 Similarity for voltage at left and right temporal sites during image repetitions: Interaction between spacing, memory, and time [ <i>n.s.</i> ]. . . . .	44
3.11 Similarity for oscillatory power at left and right parietal sites during image repetitions: Interaction between spacing, memory, and time [ $p = .0535$ ]. . . . .	45
4.1 N1 to words at electrode 58 (T5) and neighbors, analyzed window 94–194 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The early negative peak is not different across spaced and massed conditions. . . . .	55

4.2 N400 to words at electrode Cz and neighbors, analyzed window 227–427 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The negative peak at 400 ms is significantly smaller for massed repetitions compared to any spaced repetition condition. . . . .	56
4.3 LPC to words at electrode T7 and neighbors, analyzed window 436–636 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The positive peak around 600 ms is significantly larger for massed compared to spaced repetitions. . . . .	57
4.4 Theta power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the two time windows (error bars are SEM). . . . .	61
4.5 Lower alpha power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the two time windows (error bars are SEM). . . . .	63
4.6 Upper alpha to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the three time windows (error bars are SEM). . . . .	64
4.7 Lower beta power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the three time windows (error bars are SEM). . . . .	65
4.8 Similarity for voltage at left and right temporal sites during image repetitions: Interaction between spacing, memory, and time [n.s.]. . . . .	68
4.9 Similarity for oscillatory power at left and right parietal sites during image repetitions: Interaction between spacing, memory, and time [n.s.]. . . . .	70

# Chapter 1

## General introduction

“You can get a good deal from rehearsal,  
If it just has the proper dispersal.  
You would just be an ass,  
To do it *en masse*,  
Your remembering would turn out much worsal.”

—Ulric Neisser,  
quoted in “Retrieval practice and the maintenance of knowledge” (Bjork, 1988)

### 1.1 Overview of the spacing effect

The Latin phrase *repetitio est mater studiorum* tells us that “repetition is the mother of learning,” and psychological research shows us that how we distribute those repetitions has an important impact on memory performance. From the beginnings of empirical research on memory it has been shown that distributed practice, with gaps between study sessions, leads to better long-term memory performance than massed practice. Ebbinghaus documented this effect and wrote, “With any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time” (Ebbinghaus, 1885/1913, p. 89). This seemingly simple result is known as the spacing effect, or the distributed practice effect, and has been the subject of extensive research over recent decades (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Cepeda et al., 2009; Delaney, Verkoeijen, & Spriegel, 2010).

The spacing effect is robust and has been demonstrated in studies that employ various memory tests including free recall, cued recall, recognition, and frequency judgments. It is found not only in laboratory studies but also in real-world training and learning settings. Research on this topic could have important practical consequences regarding how information is presented in ap-

plied settings such as classrooms (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Khajah, Lindsey, & Mozer, 2014) and how students are instructed to study on their own, though this chance to enhance memory in applied settings depends on effectively utilizing the knowledge gained from research (Dempster, 1988; Pashler, Rohrer, Cepeda, & Carpenter, 2007).

The reasons why this almost ubiquitous effect occurs are still debated and levels of prominence for different theories have changed over the years. Three theories have come to dominate, each of which has been supported by behavioral results and different verbal and mathematical models of learning. These three theories are typically known as deficient processing, contextual variability, and study-phase retrieval.

Although behavioral studies and modeling efforts have done a commendable job of investigating the spacing effect during decades of research, few publications have investigated its neural correlates. Because these theories emphasize several different processes underlying the spacing effect, it is proposed that each may be supported or challenged by examining how particular patterns of neural activity during spaced and massed learning lead to different memory outcomes, thereby revealing the true mechanisms behind why distributed practice is so effective. Note that the theories are not mutually exclusive.

Ancillary to this main goal, even if decisive conclusions cannot be drawn regarding what is truly at the root of spacing effects, the present research describes the neural correlates of the spacing effect that align with each hypothesis. If existing mathematical models that account for the spacing effect are accurate (reviewed later), this will likely involve interactions between the theories. Thus, we can gain a better understanding of the neural processes involved during spaced and massed learning episodes as recorded in the electroencephalogram (EEG). The experiments and analyses used EEG to infer the relative involvement of various cognitive factors, thereby providing evidence relevant to assessing the theories.

## 1.2 Background for evaluated hypotheses

### 1.2.1 Deficient Processing

The deficient processing theory focuses on the cognitive processes that are active during encoding. This hypothesis predicts that when an item is repeated in a massed fashion (immediate repetition), because it is familiar and already in working memory there is a greater decrease in attention or encoding effort to the second presentation than there would be to a novel item or spaced repetition. According to Greene (1989), this could manifest as less rehearsal relative to a spaced repetition due to the stimulus's level of familiarity, and not to an experimental variable such as a difference in the relative amount of rehearsal time available for massed and spaced items. This could happen either voluntarily through the explicit control of attention (Greene, 1989), or involuntarily via a habituation mechanism like neural repetition suppression (Callan & Schweighofer, 2010; Hintzman, 1974; Van Strien, Verkoeijen, Van der Meer, & Franken, 2007; Wagner, Maril, & Schacter, 2000; Xue et al., 2011) or via short-term priming and transfer-appropriate processing (Challis, 1993; Mammarella, Russo, & Avons, 2002; Russo, Parkin, Taylor, & Wilks, 1998).

Even though this theory has been supported by recent research examining the neural basis of the spacing effect (specifically in relation to neural repetition suppression; Callan & Schweighofer, 2010; Van Strien et al., 2007; Xue et al., 2011), the deficient processing hypothesis cannot completely explain the spacing effect for a few reasons. The first is because of its reliance on working memory, such that it only seems to be a viable explanation at short timescales. Past research has shown that the spacing effect occurs across multiple days, weeks, and even months (Cepeda et al., 2006, 2009). It might be possible to test whether different extents of deficient processing occur using a gradation of spaced inter-study lags (e.g., deficient processing should decrease as lag increases), but other mechanisms are still needed.

Second, Delaney et al. (2010) called deficient processing an “impostor” spacing effect. Massed items are put at a disadvantage rather than giving the spaced items an advantage, and so this theory accounts for the spacing effect with a relative difference between the two. Since we want to

know about a true spacing effect (i.e., an advantage for spaced items), we must assume that other mechanisms are at play.

Finally, decades of theoretical and mathematical accounts of human memory have stressed the importance of the context that accompanies studied material in influencing memory performance, and the deficient processing theory does not consider this aspect. (Include references or another sentence about context?) In summary, the deficient processing effect is still interesting to investigate due to the recent attention it has been given in functional magnetic resonance imaging (fMRI) publications regarding neural repetition suppression, but other theories are needed because it cannot completely account for the spacing effect.

### 1.2.2 Contextual Variability

Contextual variability theory assumes that item study presentations are associated with a slowly drifting background context where the context gets encoded in a memory trace along with the item (Bower, 1972; Estes, 1955; Melton, 1970). The set of features that constitute context is not completely consistent across different theories and models of memory, but most accounts agree to a basic extent. Context typically includes the incidental background stimuli that are present during encoding (e.g., the experimental backdrop and non-relevant stimuli) as well as the internal state of the learner; it drifts or fluctuates as time passes (Glenberg, 1979; Lohnas, Polyn, & Kahana, 2011; Malmberg & Shiffrin, 2005; Raaijmakers, 2003). The contextual state can be impressionable, influenced by recent experiences (e.g., the other items in a list) (Howard & Kahana, 2002; Sederberg, Howard, & Kahana, 2008). This means that context can change depending on the information that is encoded or retrieved from memory. This thesis takes the view that the contextual state fluctuates over time and is influenced by the contents of memory, as this is how spacing effects are typically accounted for in mathematical models of memory.

Because episodic context drifts with time, the context encoded with repeated study events will differ more as inter-study lag increases (Estes, 1955). This was first proposed as a reason for spacing effect by Melton (1967) and was integrated into more elaborate theories (Bower, 1972;

Glenberg, 1976, 1979). Subsequent retrieval of an item during test depends, at least partially, on the similarity between the study and test contexts (in line with the encoding specificity principle Tulving & Thomson, 1973). Consequently, spaced items are recalled better due to the higher probability that the contextual state at test will match that of the spaced repetitions compared to the less variable (or nearly identical) massed repetitions; essentially, there are more retrieval cues for spaced items. To account for decreased performance at especially long study repetition lags, Glenberg (1976) proposed that the test context would not overlap with the first presentation and memory would rely solely on the second study presentation.

As a side note, attempts to deliberately vary the context encoded with massed and spaced repetitions (e.g., changing the paired associate, using homonyms, varying the level of processing), have typically not produced spacing effects (e.g., Glanzer & Duarte, 1971; Hintzman, Summers, & Block, 1975) (but see Braun & Rubin, 1998, Experiment 3; Malmberg & Shiffrin, 2005). This is typically accounted for by assuming that when context is too variable the item is no longer encoded as a repetition. Therefore, attributing spacing effects to temporocontextual drift rather than strictly defined contextual cues seems more promising (Greene, 1989).

To extend contextual variability's predictions, it seems natural that contextual drift should apply to the learner's ongoing neural activity. Importantly, this has been considered with respect to memory models (e.g., Sederberg et al., 2008) and demonstrated in analyses of intracranial EEG in rats (Manns, Howard, & Eichenbaum, 2007) and humans (Manning, Polyn, Baltuch, Litt, & Kahana, 2011). Therefore, this theory can be assessed by measuring the neural similarity between study repetitions as well as during test trials.

### 1.2.3 Study-Phase Retrieval

Study-phase retrieval proposes that a stimulus repetition can induce a retrieval attempt for the memories of its earlier presentation(s). An item and its contextual information would be retrieved, and retrieval during study is important for improving subsequent memory under this theory (Greene, 1989; Thios & D'Agostino, 1976). Surely massed items will be recognized as being

repeated as well, so why do the spaced items benefit more (the crux of the spacing effect)? As with contextual variability, context varies more across spaced repetitions (Melton, 1967). However, under study-phase retrieval a repetition is assimilated into the existing memory trace, thereby making the trace stronger (Raaijmakers, 2003). This provides an addition set of retrieval cues for spaced items, including the temporal relation between the events (Greene, 1989). Many descriptions of this theory require that the learner recognize the repeated item in order to update the trace and subsequently benefit from spacing (Hintzman & Block, 1973; Hintzman et al., 1975; Johnston & Uhl, 1976; Raaijmakers, 2003). Otherwise a new memory trace may be stored for the repetition. Thus, if study-phase retrieval occurs, this supersedes the contextual variability hypothesis, which does not posit any retrieval attempts.

As lag increases between repetitions it may become more difficult to retrieve the prior study event, but if a longer lag item is retrieved then the memory trace is strengthened to a greater extent (Delaney et al., 2010). This is essentially a description of a lag effect (Glenberg, 1979; Greene, 1989) and has been demonstrated in the literature (e.g., Kahana & Howard, 2005). Some of the theories and models described below account for this, for example more contextual cues are stored at longer repetition lags, and the mechanisms involved are part of the investigations performed in this thesis. Notably, this retrieval difficulty idea is in line with research on the testing effect concerning the “desirable difficulty” of a test. For example, Karpicke and Roediger (2007) showed that a more difficult retrieval due to repeating an item at a longer delay promoted long-term retention (here, the repetition was a test).

This theory brings up the question of how the memories are stored. Is a recognized repetition stored as a new trace, or is the original trace strengthened? As alluded to above, Raaijmakers (2003) explained that a repetition should strengthen (and add to) the initial stimulus trace. If each repetition was instead stored as a separate trace, the first trace will decay at longer lags (effectively a long retention interval) and subsequent memory may rely solely on the second trace, a potential issue for the contextual variability theory. However, this is not the typical result of a spacing manipulation, except at very long lags as Glenberg (1976) explained. Updating an existing

trace with new information is implemented in other models as well (e.g., Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009).

If a stimulus repetition brings to mind its prior occurrence(s), study-phase retrieval theory predicts contextual reinstatement at a neural level (Howard & Kahana, 2002). Contextual reinstatement occurs when the memory of a prior episode is reactivated. Say list  $L$  has  $n$  presentations, some of which are repetitions, and  $L_1$  and  $L_{n-1}$  are the same item presented at two different points in time with a number of intervening items (e.g.,  $n > 10$ ). Study-phase retrieval says that the neural activity during  $L_{n-1}$  will be more similar to its initial presentation  $L_1$  than to another item presented at nearby list position  $L_{n-2}$ . This is different than what would be expected under the contextual variability theory where context should simply drift:  $L_{n-1}$  would be more similar to  $L_{n-2}$  than to  $L_1$ . Turk-Browne, Simon, and Sederberg (2012) demonstrated this using fMRI by showing that the same visual scene presented on different lists but preceded by the same stimulus (i.e., the same context) was found to evoke more similar neural activity compared to when it was preceded by different stimuli (i.e., different contexts). This can support the contextual reinstatement aspect of study-phase retrieval because the presentation of the same information puts the brain in similar states. Additionally, Lohnas and Kahana (2014) implemented reinstatement in a model (described below) that accounts for spacing effects in free recall experiments.

Contextual reinstatement is also thought to occur during a memory test. Some memory models assume that the current contextual state is used as a cue during retrieval attempts, especially during a free recall test without experimenter-supplied cues. When an item is recalled, the present context is updated with that item's associated context (Howard & Kahana, 2002; Sederberg et al., 2008). This assumption helps account for particular patterns of recall in experimental evidence (Kahana, 1996; Lohnas & Kahana, 2014). Additionally, evidence for similar neural activity during study and recall of a given stimulus shows that this kind of contextual reinstatement occurs in the brain (Manning et al., 2011; Polyn & Kahana, 2008; Xue et al., 2010). Thus, under study-phase retrieval we would expect both study repetitions and seeing the item during a subsequent test to reinstate prior context.

#### **1.2.4 Hypothesis interactions**

It is important to note that these theories are not mutually exclusive or necessarily competing, and it would be difficult, if not impossible, to test each independently. In fact, they can work well together, and the consensus in the literature is that a hybrid account is needed to explain spacing effects (Delaney et al., 2010; Greene, 1989; Lohnas & Kahana, 2014; Lohnas et al., 2011; Raaijmakers, 2003). Most hybrid accounts agree that each presentation is encoded with drifting context (contextual variability) and that an item repetition is assimilated with prior occurrences (study-phase retrieval). This combination seems necessary to account for the effect when fitting models to empirical data (e.g., Lohnas & Kahana, 2014; Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009; Pavlik & Anderson, 2005).

### **1.3 Formal models accounting for the spacing effect**

An important way to constrain explanations of the spacing effect is with the mechanisms implemented in computational models that can account for patterns of human performance. Most of these models contain interacting mechanisms that involve more than one of the hypotheses described here.

Raaijmakers (2003) made an influential model based on the Search of Associative Memory model (SAM; Mensink & Raaijmakers, 1989) using the spacing effect theory proposed by Glenberg (1979). It accounts for the spacing effect in cued recall using contextual variability and study-phase retrieval mechanisms.

An activation-based account of the spacing effect was implemented by Pavlik and Anderson (2005) in ACT-R, and can explain the effect at multiple timescales, which the SAM-based model cannot. Here, all repeated items receive a strength increment, but the rate of decay for the resulting trace is positively correlated with the level of activation for that item at the time of its repetition. This leads to a spacing effect that fits their behavioral data well. The quick decay for a massed item is reminiscent of deficient processing, but the mechanism is not explicitly defined in this way.

The Multiscale Context Model (MCM; Mozer et al., 2009) can also account for the spacing effect at multiple timescales. It makes impressively accurate predictions at various inter-study lags and retention intervals, as well as for different study materials, using relatively few parameters. It implements contextual variability and study-phase retrieval (“retrieval-dependent update”) in a method similar to the SAM model for storing and retrieving context and item information. Additionally, it uses a mechanism similar to the ACT-R model’s decay to predict a forgetting function.

The Context Maintenance and Retrieval model (CMR; Lohnas & Kahana, 2014; Polyn et al., 2009) can account for the spacing effect in free recall. As its name implies, contextual variability is an essential aspect for modeling the encoding of context as it fluctuates, and a study-phase retrieval mechanism helps it reinstate previous contextual states. These two aspects allow it to capture patterns of spacing effect results in a paradigm that the other models were not designed to fit.

Overall, contextual variability and study-phase retrieval are clearly important theories, as these mechanisms are central to successful models that capture the spacing effect. Even though none of these models use deficient processing, it should be given attention due to its recent popularity in empirical investigations involving neural recordings.

#### 1.4 EEG studies of the spacing effect

Van Strien et al. (2007) published the only study investigating the spacing effect using event-related potential (ERP) and time-frequency analyses, which was based on stimulus repetition research. They used a continuous recognition paradigm (repeatedly judge word presentations as being new or old) followed by a surprise recall test. Unfortunately, some of their EEG effect interpretations do not agree with episodic memory research and instead are explained as being specific to working memory paradigms, though it seems unclear whether theirs is a working memory task (spaced repetitions were 6 items/9 sec apart). This does not seem ideal to base new episodic memory research on.

Although the behavioral results of Van Strien et al. (2007) indicate higher recall for spaced *vs.* massed repetitions, they did not analyze neural data for an interaction between spacing and memory which seems like an important factor to investigate in a study of the spacing effect. Also, their use of continuous recognition may have confounded their interpretation of the data because this paradigm induces a testing effect (every trial is an old/new test). Behavior may differ qualitatively compared to simply studying and encoding stimuli at spaced and massed intervals (Delaney et al., 2010, p. 91). Thus, further research is needed to understand the EEG patterns and cognitive processes that underlie the spacing effect. Despite these shortcomings, their results can be used as a basis for future experiments to make effect predictions when modulated by subsequent memory, an analysis that has not been previously explored using neural data.

## **Chapter 2**

### **Dependent measures**

The terms subsequent memory and subsequent retrieval were mentioned earlier. To be clear on these and related terms, subsequent memory effects (SMEs) show differences during encoding that are contingent upon later memory performance, and examining these differences sheds light on why some things can be remembered later while others are forgotten. Additionally, the retention interval between the final study event and the test event will also have an impact on the optimal lag (Cepeda et al., 2006; Glenberg, 1976, 1977, 1979); however, optimal lag will not be considered here.

Because the spacing effect predicts a long-term memory advantage for spaced compared to massed study repetitions, it seems likely that there will be differences between spaced and massed repetitions that interact with subsequent memory. However, it also seems likely that when encoding is successful for both spaced and massed repetitions, the active memory mechanisms should be qualitatively similar. There may be a quantitative difference (the difference is in the degree of processing), but it remains possible that memory will not be critical to determining why the spacing effect occurs; many factors besides encoding mechanisms can influence subsequent memory. Although we will investigate neural patterns known to be associated with memory encoding and retrieval, other cognitive factors must be considered. For example, attention and semantic processing likely influence the spacing effect. Effects of spacing for repetition events without a subsequent memory interaction can also reveal processing differences between spaced and massed items.

## 2.1 Neural data

EEG recordings provide a fine-grained time course of the electrical activity of neurons, on the order of milliseconds, which is important for determining when neural processes occur with respect to behavioral responses. EEG recorded across the scalp, as in the present experiments, represents the combined activity of millions of neurons. For analysis, EEG is typically averaged across brief epochs that are time-locked to events such as stimulus presentations, and these event-related potentials (ERPs) are compared between conditions to show differences in voltage deflections, known as ERP components. Particular ERP components, which dissociate in time (relative to stimulus presentation) and space (at particular scalp electrodes), have come to be associated with cognitive processes.

Similarly, decomposing EEG into time-frequency measures using methods such as wavelet transforms (think Fourier transforms, but estimating the amplitude, or power, of activity at a particular frequency across time) provides a more nuanced (and higher-dimensional) representation of neural activity. This allows for an investigation of the oscillatory dynamics of neural networks. Oscillatory power is used to measure both local synchronous activity and long-range communication between brain regions, and different frequency bands have also come to be associated with particular cognitive processes. Power can be measured as a change relative to a pre-stimulus baseline, so when a frequency band increases in power (synchronizes) or decreases in power (desynchronizes), this means that the amplitude in that frequency band was higher or lower compared to the baseline.

## 2.2 Effects of attention and semantic processing

Assessing how attentional and semantic processes are modulated by spaced and massed presentation is central to the deficient processing theory. The present understanding is that an involuntary decrease in stimulus processing occurs for repetitions of recently encountered (massed) items, while this does not occur for spaced repetitions.

Under the deficient processing hypothesis, massed repetitions should involve decreases in

attentional processes. Because early ERP components have been related to attentional processing (Luck, Woodman, & Vogel, 2000), these may show spacing effects. The visual N1 ERP component (early negative occipitoparietal peak) shows effects of selective attention (Klimesch et al., 2004) particularly during discrimination tasks (Vogel & Luck, 2000), and is related to increased visual analysis (Curran, Tanaka, & Weiskopf, 2002). It also shows effects of lexical and semantic processing (Proverbio & Adorni, 2009). It does not typically show subsequent memory effects (e.g., Curran et al., 2002; Duarte, Ranganath, Trujillo, & Knight, 2006; Duarte, Ranganath, Winward, Hayward, & Knight, 2004), and may act like an early attentional gating mechanism and attenuate for massed repetitions.

The analysis of neural oscillations has become an important tool for cognitive neuroscientists (for reviews, see Hanslmayr & Staudigl, 2014; Nyhus & Curran, 2010). Oscillatory desynchronization (decrease in power) in the lower alpha band (8–10 Hz; widespread) is related to increased attention (Klimesch, 1999; Klimesch, Sauseng, & Hanslmayr, 2007; Pfurtscheller & Lopes da Silva, 1999), though the functional interpretation of lower alpha is generally less clear than for other bands (Klimesch et al., 2007). Because deficient processing occurs for massed items, we would expect lower alpha to show more power (less desynchronization) for massed than spaced.

Another idea tied to deficient processing is priming, specifically semantic (Challis, 1993) and perceptual priming (Mammarella et al., 2002; Russo et al., 1998). A primed and still-active representation during a massed repetition will require less activation of semantic or perceptual features compared to spaced repetitions, and thus less processing will occur. The N400 ERP component (negative central peak) is affected by semantic processing and priming (for a review, see Kutas & Federmeier, 2011). Performing a task involving semantic analysis will produce a more negative N400, linking it to the activation of semantic information (Kutas & Hillyard, 1980). The component is also linked to semantic anomalies, but in general, the N400 is thought to reflect the dynamic construction of semantic meaning as part of a larger distributed system (as opposed to being a direct measure of accessing the meaning of a stimulus).

N400 effects are seen in repetition paradigms where the component attenuates to repetition

events compared to initial presentations (Kim, Kim, & Kwon, 2001; Olichney et al., 2000; Van Strien et al., 2007); voltages becomes more negative as lag between repetitions increases. The amount the component attenuates to a stimulus repetition or to a target stimulus following a semantic priming stimulus reveals “how much of the information normally elicited by that stimulus is already active” (Kutas & Federmeier, 2011, p. 23). Spaced repetitions elicit an N400 (representing the activation of semantic processing) whereas this does not occur for massed repetitions since the information is primed and in working memory (Van Strien et al., 2007). This repetition attenuation should be a factor under deficient processing since failing to receive additional semantic processing events is one way subsequent memory could be worse for massed items.

For oscillatory effects, upper alpha (11–12 Hz; posterior) desynchronization (decrease in power) is related to the reactivation of semantic information from long-term memory (Klimesch, 1999; Klimesch, Schack, & Sauseng, 2005). Decreases in power in the lower beta band (13–21 Hz; central and temporal) are similarly associated with the semantic processing of to-be-remembered items (Fellner, Bäuml, & Hanslmayr, 2013; Hanslmayr, Staudigl, & Fellner, 2012; Hanslmayr et al., 2011). If these bands reflect retrieving information from memory, we would expect to see less power in both for massed compared to spaced because the information is already primed and exists in working memory. However, if the power decreases reflect the processing of semantic information after retrieval, both massed and spaced should show desynchronization, and massed repetitions may actually show an earlier onset because the information can be accessed faster.

### 2.3 Effects of memory retrieval and encoding

Memory processes should also show effects of study repetition lag, as long-term memory performance is the critical measurement of the spacing effect. This area of analysis can address all three theories, but provides a particularly important assessment of study-phase retrieval.

The parietal old/new effect (late positive peak, also sometimes called the late positive component or LPC) is an ERP component that indexes conscious recollection. Its amplitude correlates with the subjective amount of retrieved information (Vilberg, Moosavi, & Rugg, 2006; Wilding,

2000; Wilding & Rugg, 1996). The LPC is also seen in repetition experiments (Olichney et al., 2000; Van Strien et al., 2007), and is linked to the conscious recognition of an item. Under deficient processing, there should be a larger effect for massed than spaced repetitions relative to new items due to the higher perceived memory strength of massed items. This falls under the idea that massed items are primed and in working memory, while spaced items are not. On the other hand, study-phase retrieval requires episodic retrieval during a spaced repetition while deficient processing and contextual variability do not. Because spaced repetitions need to be retrieved from long-term memory but massed do not, if the LPC indexes retrieval then study-phase retrieval theory might predict a larger effect for spaced repetitions than for massed. Based on the prior stimulus repetition research showing a larger effect for massed items, we expect the deficient processing explanation to occur.

Theta power (4–7 Hz; frontal, temporal, and parietal) is related to memory formation and retrieval, particularly in medial temporal lobe regions (Klimesch, 1999; Klimesch, Doppelmayr, Russegger, & Pachinger, 1996; Klimesch et al., 2006; Long, Burke, & Kahana, 2013; Sederberg, Kahana, Howard, Donner, & Madsen, 2003; for a review, see Mitchell, McNaughton, Flanagan, & Kirk, 2008). It is also thought to reflect item–context binding (Hanslmayr, Spitzer, & Bäuml, 2009; Hanslmayr et al., 2011; Staudigl & Hanslmayr, 2013; Summerfield & Mangels, 2005). Theta power typically increases during encoding for subsequently remembered items (a positive SME), though negative SMEs have been demonstrated (e.g., Burke et al., 2013; Lega, Jacobs, & Kahana, 2011; for a review, see Hanslmayr & Staudigl, 2014). Theta power should increase for spaced compared to massed repetitions, but for different reasons under each hypotheses. Under deficient processing, theta would decrease for massed repetitions simply because the item is not being retrieved or re-encoded well. Under both contextual variability and study-phase retrieval, theta would increase for spaced repetitions because the intervening context (and the prior presentation in the latter case) is also encoded (new information, item–context binding). However, there is a difference between these theories: given that a repetition is properly re-encoded, in comparison to the initially encoded memory (Presentation 1), the “contents” of the two encoding events will be more similar under

study-phase retrieval and more variable under contextual. This brings us to the next topic.

## **2.4 Memory reinstatement and contextual variability assessed via neural similarity**

We do not just want to know about attention, the extent of semantic processing, and memory strength; we also want to investigate the contents of memory and whether it evolves across study repetitions. We can address this by assessing whether the similarity of neural activity for an initial presentation and repetition is correlated with subsequent memory performance, and whether this similarity is modulated by spacing. These assessments will bear on both study-phase retrieval and contextual variability. This can be tested both as a main effect of similarity (“Does greater similarity or greater variability in neural activations during encoding lead to better subsequent memory?”) and as an interaction with spacing (“Do spaced and massed repetitions benefit from greater similarity or greater variability during encoding? Are there differences?”).

Although each theory predicts better subsequent memory performance (recognition, recall, etc.) for spaced compared to massed learning (except at very long retention intervals) and naturally includes encoding mechanisms as reasons for the effect, not all implicate processing at the time of test as having an impact on memory performance. The similarity between an item’s study repetitions and its subsequent memory test event is a meaningful comparison for contextual variability and study-phase retrieval. For a spaced item, encoding variability would predict that the test is similar to only one of the study presentations, while it would likely be similar to both under study-phase retrieval. Comparisons for massed items would be more difficult to interpret since context would not have drifted much.

There are different ways to measure the similarity of neural activity between individual events. For example, Representational Similarity Analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) has been used in the fMRI literature recently (e.g., Xue et al., 2010) but has hardly been applied to EEG (e.g., Groen, Ghebreab, Lamme, & Scholte, 2012; Su, Fonteneau, Marslen-Wilson, & Kriegeskorte, 2012) and has never been used in a memory experiment. Other multivariate analyses

have been used in electrocorticography (ECoG, or intracranial EEG) in relation to contextual drift (e.g., Manning et al., 2011), and would be suitable here if used appropriately. Regardless of the method used, a measure of neural similarity will help assess the nature of neural representations detectable at the scalp. A challenge of this analysis is that it is difficult to know whether similarity measures are comparing item features, contextual features, the currently engaged mechanisms, or some combination of these; this analysis would benefit from separating item and context features in neural representations.

## 2.5 Summary

The overarching goal of this thesis is to assess the major hypotheses for why the spacing effect occurs by examining data from experiments that capture this effect. We recognize that we are relying on reverse inference to test these psychological theories of the spacing effect (Poldrack, 2006; Poldrack & Wagner, 2004): patterns of neural activity are used to make assumptions about active cognitive processes, as related to prior research. Since there is so little research on the EEG correlates of the spacing effect, reverse inference provides an initial direction for our analyses. To briefly recap the factors that likely influence each theory, deficient processing emphasizes attention and semantic processing, contextual variability emphasizes contextual drift and study–test matching, and study–phase retrieval emphasizes episodic retrieval of the prior presentation during a repetition. We expect that both theory- and data-driven analyses of ERPs, oscillatory power, single trials, and study–study and study–test neural similarity will support or challenge the theories, and will provide insight into the cognitive underpinnings of the spacing effect.

Experiment 1 involved a paired associates memorization task in which participants studied word–image pairs at two points in time in either a spaced or a massed fashion. After a brief distractor task, a cued recall test was given where participants were required to remember the word originally paired with each image. ERP and oscillatory effects were used to assess how attention, semantic processing, and memory retrieval and encoding mechanisms operate under spaced and massed learning and as modulated by subsequent memory. Neural similarity was performed via

dimensionality reduction techniques to attempt to understand memory content and contextual reinstatement. Experiment 2 was an extension run using additional study repetition lags and used a similar design as well as similar analyses.

## **Chapter 3**

### **Experiment 1**

This experiment provides an investigation into how neural activity during paired associate learning changes as a function of massed or spaced practice.

#### **3.1 Method**

##### **3.1.1 Participants**

Thirty-seven University of Colorado Boulder undergraduates participated in the experiment for either course credit ( $n = 17$ ) or payment of \$15 per hour ( $n = 20$ ) (ages 18–26,  $M = 20.5$ ; 13 female). All participants were right-handed native-English speakers and had normal or corrected-to-normal vision. Informed consent was obtained from each participant, and the study conformed to the Institutional Review Board guidelines.

##### **3.1.2 Materials**

The experimental stimuli were randomly selected from 1521 common nouns (the PEERS word pool<sup>1</sup>) and 832 images (two categories: 371 face images and 461 living room scenes). Face images were shoulder-up photographs taken in front of an off-white background with the center of the face generally in the center of the image (Phillips, Moon, Rizvi, & Rauss, 2000). Scene images were photographs taken from the SUN image database within the “Living Room” category (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). Words were presented in Courier font (size 24) and all

---

<sup>1</sup> <http://memory.psych.upenn.edu>

images were cropped to be the same size ( $480 \times 320$  pixels). Stimuli were presented on a 17-in flat-panel display with a resolution of  $1024 \times 768$  (60 Hz frame rate) placed 1 m in front of the participants. All portions of the display not occupied by stimuli or text were filled with gray pixels.

The experiment was programmed in MATLAB (versions R2012b and R2014a; The MathWorks, Inc., Natick, MA) using our experimental framework<sup>2</sup> and was presented using Psychtoolbox (Brainard, 1997).

### **3.1.3 Design**

Experiment 1 consisted of six blocks of four experimental phases (Figure 3.1): exposure, study, distractor, test. The session, including application of the electrode net and running in the task, lasted approximately 2.5 hours. Stimuli were randomly shuffled prior to creating the list for each phase at the beginning of the session. The study phase contained the conditions that were manipulated within subjects, namely the viewing of spaced and massed paired associates.

### **3.1.4 Procedure**

An electrode net was applied to each participant’s head, and the session began with a shortened practice version of the experiment to familiarize participants with the study and test procedures (two spaced, two massed, and two single presentation items, with a lag of 4 items between spaced presentations; two new images were included at test).

In the exposure phase, participants viewed the 50 images (half faces and half scenes, randomly intermixed) that they would subsequently see on the study list and rated each on a four-point “appealing” scale: very appealing, somewhat appealing, somewhat unappealing, and very unappealing. Only the images from the upcoming study period were presented; the words were not shown. The keys D, F, J, and K were used to make the response, and the scale-to-keyboard mapping was counterbalanced across participants. An image denoting the key-response mapping was shown at the bottom of the screen at all times, but participants were encouraged to memorize the keys so

---

<sup>2</sup> <https://github.com/warmlogic/expertTrain>

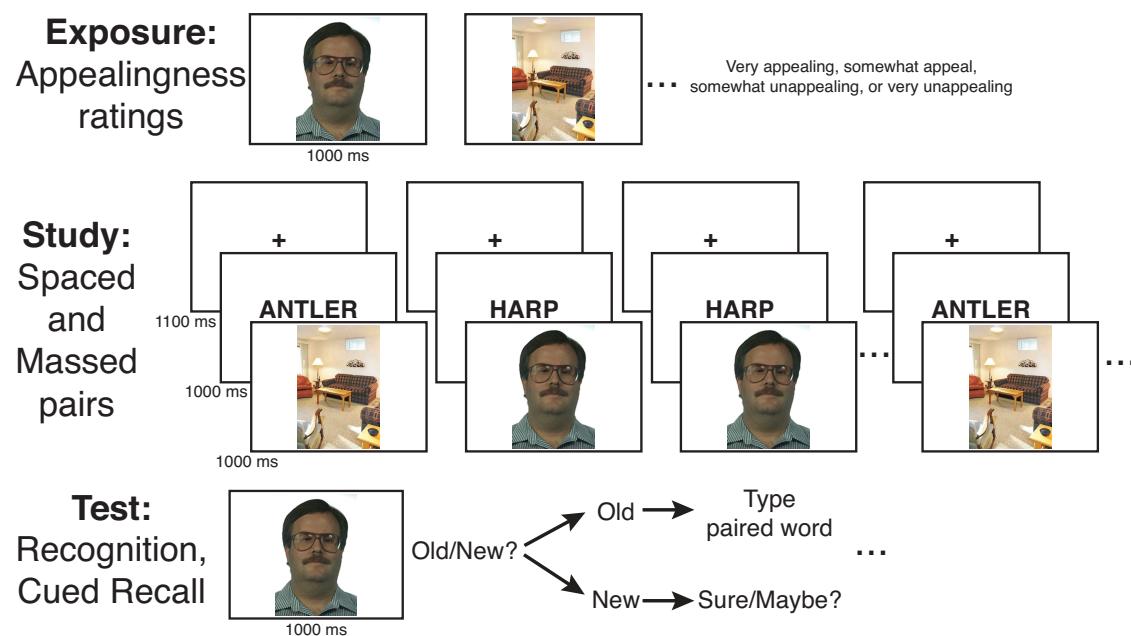


Figure 3.1: Experiment 1: Exposure, study, and test phases

they could keep their eyes on the fixation cross at the center of the screen. On each trial, a cross was shown for 1.0–1.2 s (jittered), then the cross and image were shown for 1.0 s, after which the cross changed to a question mark prompting participants to make a response. Participants were allowed to respond during the initial 1.0 s image presentation; if this occurred, the image stayed on screen for a total of 1.0 s. If they waited longer than 1.0 s and the cross changed to a question mark, the image remained on screen until a response was made or 3.0 s passed. No more than three images from the same category could occur in a row. This phase lasted approximately 3 min.

In the study phase, participants viewed 50 word–image pairs and were asked to think of a relationship between them or to make up a story pairing them. They were told that a subsequent test would require them to remember the word associated with each picture, but they were not told that some pairs are repeated. For each of the two image categories there were seven two-presentation spaced pairs, seven two-presentation massed pairs, seven pairs presented only once, and four additional single-presentation buffers (two at the beginning of the list and two at the end). Only images from the spaced and massed pairs were included on the test list. Spaced items were presented at a lag of 12 (12 intervening pairs between presentations 1 and 2), and massed items were presented at a lag of zero. On each trial, a fixation cross appeared for 1.0–1.2 s (jittered), then the word was presented first for 1.0 s followed immediately by the image for 1.0 s. No more than three images from the same category could occur in a row, and no more than two trials with the same lag (including single-presentation pairs) could occur in a row. Each study phase lasted approximately 5 min.

In the distractor phase, participants answered simple math problems of the format  $A+B+C=?$  for either 2 min or until they answered 60 problems, whichever came first. They typed their responses with the keyboard. Different tones occurred for correct and incorrect answers, and mean accuracy and response time was reported to the participant at the end of the phase.

Finally, in the test phase, participants performed recognition and cued recall tasks. Twenty-eight old images (seven spaced and seven massed from each category) and 14 new images (seven lures from each category) were mixed together and presented one at a time, at which point participants

made two responses. First, they had to decide whether the image was studied earlier (“old”) or had not been seen before (“new”) using the F and J keys (counterbalanced). If they answered “old”, they saw ??????? below the image and had to type the word previously paired with the image; they could pass if they could not remember the word. If they answered “new”, they either said that they were “sure” it was new or it was “maybe” new using the F and J keys (counterbalanced); this confidence judgment was used so the same number of responses occurred for both “old” and “new” items. An image showing the key–response mapping was shown at the bottom of the screen when appropriate, but participants were encouraged to memorize the keys so they could keep their eyes fixated on the cross. On each trial, a fixation cross appeared for 1.0–1.2 s (jittered) and the image was shown for 1.0 s, at which point the cross turned to a question mark and participants were asked to make their initial recognition response. With lures mixed in, no more than four images from the same category could occur in a row. Importantly, test images were presented in a sequence similar to the study order. To construct the test list, the positions of the second presentations of study stimuli were divided into five contiguous groups and each group was randomly shuffled. This was done to approximately preserve a similar amount of time between the second presentation and the test across all “old” stimuli. Each test phase lasted approximately 4 to 5 min.

### **3.1.5      Electrophysiological recordings and data processing**

A 128-channel HydroCel Geodesic Sensor Net<sup>TM</sup> (GSN 200, v. 2.1; Tucker, 1993) was used to measure the EEG at the scalp using a central vertex reference (Cz) with a sampling rate of 250 Hz and a low-pass hardware filter at 100 Hz (see Figure 3.2). The net was connected to an AC-coupled, high-input impedance amplifier (300 MΩ, Net Amps<sup>TM</sup>; Electrical Geodesics, Inc., Eugene, OR) and recordings were made using the Net Station application. The electrodes were adjusted until impedance measurements were less than 40 kΩ.

All data processing steps and analyses were done in MATLAB using in-house scripts<sup>3</sup> and the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). A high-pass filter at 0.1 Hz,

---

<sup>3</sup> <https://github.com/warmlogic/mat-mvm>

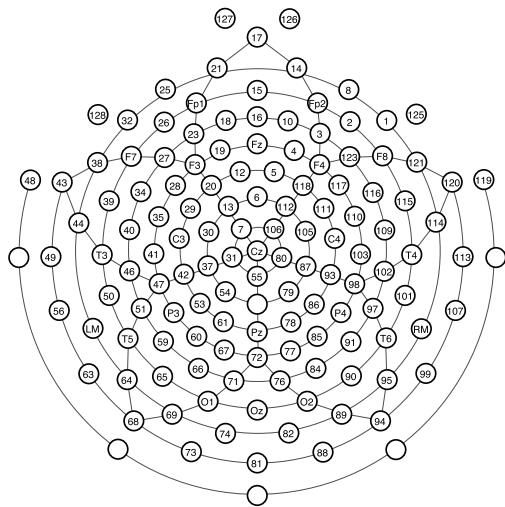


Figure 3.2: The 128-channel HydroCel Geodesic Sensor Net<sup>TM</sup> used to measure the EEG. In this top-down schematic the participant's nose is toward the front.

low-pass filter at 100 Hz, and a notch filter from 59–61 Hz were applied to the data. Study and test trials were epoched into 3000 ms segments, 1 s before the onset of each stimulus and 2 s after. Artifact detection was used to reject particularly noisy epochs, as well as those that exceed an amplitude of  $\pm 100 \mu\text{V}$ . The data were referenced to the average of all channels and individual trials were baseline corrected relative to -200 ms to 0 ms.

We would not expect differences in neural activity between the initial presentations of spaced and massed items, as the locus of the spacing effect occurs after this point. There may exist subsequent memory effects here, though it would be difficult to know whether neural activity during the first presentation is the reason for this effect (e.g., perhaps a subsequently remembered item was encoded poorly on the first presentation and very well on the second presentation). ERP analyses included these initial presentations in analyses as a baseline, but time–frequency analyses focused on the second presentation (the repetition) and show the single-presentation stimuli in data plots. The repetition is therefore the critical trial type to analyze, as this is where the massed and spaced manipulation occurs, and plots will focus on the repetition. The first presentation is included in the analysis to see whether activity changes differently across presentations for massed and spaced items (the ANOVA factor of presentation). To foreshadow some results, there were no differences between first presentations of massed and spaced items in ERP analyses; these trials were not included in time–frequency analyses.

### 3.2 Results

Thirty-one participants were included in behavioral analyses. Fourteen participants were excluded from ERP and time–frequency analyses because they either did not perform the task properly ( $n = 1$ , completely excluded) or had fewer than 10 artifact-free trials in any of the main trial conditions ( $n = 13$ ; four had extremely low trial counts and were also excluded from behavioral analyses), leaving twenty-three participants in EEG analyses. Similarity analyses included the twenty-eight participants who had six or more artifact-free pairs of initial presentation–repetition image trials.

All analyses contingent on subsequent memory are split by whether words were recalled or forgotten after the image was correctly recognized as being old. Significant results are reported, and unreported results can be assumed to not be significant. When an ANOVA contains a factor with more than two levels, the reported values are adjusted for violations of assumptions of sphericity using the Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959) even if the factors did not violate Mauchly's test of sphericity.

### 3.2.1 Behavioral results

On average, scenes were rated as more appealing compared to faces ( $M = 2.99$  vs  $M = 2.01$ ) [ $t(30) = 9.13, p = 3.6e^{-10}$ ]. The average response times during the exposure/rating phase was faster for faces (1027 ms) than scenes (1115 ms) [ $t(30) = 3.62, p = .0011$ ].

For the test phase, a two-way repeated measures ANOVA was run on image recognition discrimination ( $d'$ ) with factors of spacing and image category. There was a main effect of spacing [ $F(1, 30) = 17.4, p = .00024, MSE = 0.0756$ ] such that spaced images ( $d' = 2.95$ ) were recognized better than massed images ( $d' = 2.74$ ), and a main effect of image category [ $F(1, 30) = 104, p = 3.05e^{-11}, MSE = 0.224$ ] such that faces ( $d' = 3.32$ ) were recognized better than scenes ( $d' = 2.37$ ). An ANOVA with the same factors was run on cued recall hit rate (for old items called “old”). There was only a main effect of spacing [ $F(1, 30) = 81.8, p = 4.51e^{-10}$ ] such that spaced words ( $M = 49.8\%$ ) were recalled better than massed words ( $M = 36.5\%$ ). Words paired with faces and scenes were recalled at the same rate (faces:  $M = 43.7\%$ ; scenes:  $M = 42.5\%$ ). Thus, there are clear spacing effects for both recognition and recall.

### 3.2.2 ERP results

ERP analyses were performed on 40 Hz low-pass filtered data using repeated measures ANOVAs; pairwise comparisons were made with  $t$ -tests. Peak electrodes and latencies for the ERP effects were found by collapsing all word presentation events together (using grand averages), finding the electrode with the peak voltage within the effect time ranges, and then locating the

peak latency using that electrode and its immediate neighbors. The peak electrodes had to show typical effect patterns, and ended up being near the electrodes used by Van Strien et al. (2007): Cz for the N400 and a parietal electrode just to the right of Pz (electrode 77) for the LPC effect. The N400 peaked at 372 ms (Figure 3.4). The LPC peaked at 596 ms (Figure 3.5). For visual N1, the electrode had to show negative peaks between 150 and 250 ms and it should have precedence in the literature. Electrode 58 (T5) peaked at 172 ms (Figure 3.3). Analyses use these peak electrodes and neighbors; words during the study phase were analyzed because ERPs for images would likely be affected due to immediately following word presentations.

The average voltage and peak latency data used in statistical tests comparing massed and spaced conditions were computed for individual subjects using the electrode locations and time windows described above. Peak latency was determined by averaging the 10 time samples with the largest voltage, and voltage was averaged across the appropriate sized time window at that peak time point. Three-way ANOVAs with factors of spacing (spaced and massed), presentation (1/new and 2/old), and subsequent memory (recalled and forgotten words; all trials were subsequent recognition hits) were performed. Single presentations items were not included because they were not tested, but presentation 1 is a close analogue to a single presentation item for the purpose of comparing our results to the literature.

Since the N400 and LPC have precedent in the repetition literature, we analyzed these components for semantic processing and memory effects, and analyzed the visual N1 for attentional effects.

If attention is modulated by spacing, early ERP components may show effects. The visual N1 typically shows effects of selective attention, making this component particularly relevant to deficient processing. A three-way ANOVA with factors of spacing, presentation, and subsequent memory was performed. A crossover pattern was borne out in the significant three-way interaction [ $F(1, 22) = 10.5, p < .005$ ] that showed spacing and subsequent memory only had effects for repetitions, not for initial presentations. Recalled spaced repetitions ( $M = -1.1 \mu\text{V}$ ) were more negative than forgotten spaced repetitions ( $M = -0.39 \mu\text{V}$ ) [ $t(22) = 2.73, p < .05$ ], recalled

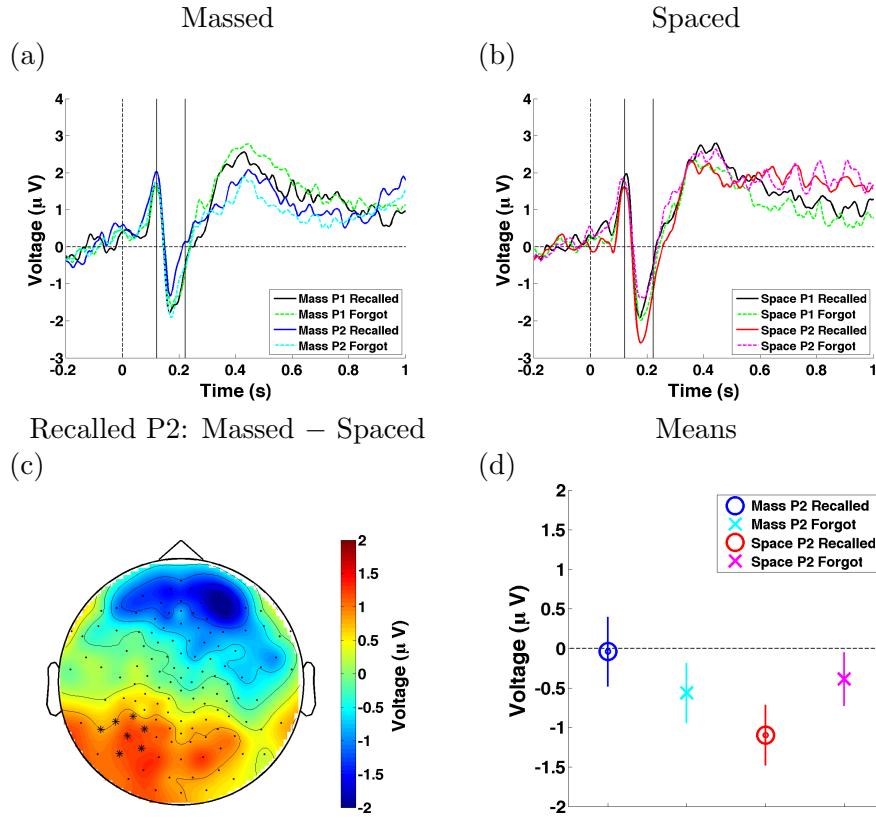


Figure 3.3: N1 to words at electrode 58 (T5) and neighbors, analyzed window 122–222 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed *vs.* spaced Presentation 2 (P2) across time and electrodes (\*); (d) analyzed means (error bars are SEM). The early negative peak is significantly larger for spaced compared to massed repetitions.

massed repetitions ( $M = -0.04 \mu\text{V}$ ) [ $t(22) = 4.37, p < .0005$ ], and (marginally) forgotten massed repetitions ( $M = -0.56 \mu\text{V}$ ) [ $t(22) = 2.039, p = .053$ ].

If the spacing effect results from differences in semantic priming and processing, the N400 should show effects. For voltage, a significant spacing  $\times$  presentation interaction showed a graded pattern [ $F(1, 22) = 12.4, p < .01$ ]: voltage becomes more negative from massed repetition to spaced repetition to first presentation [ $p < .01$ ]. Contributing to this were main effects of spacing [ $F(1, 22) = 9.65, p = .00514$ ] and presentation [ $F(1, 22) = 30.8, p = 1.42e^{-5}$ ]; spaced items were more negative, and repetitions were less negative. There was also a main effect of memory [ $F(1, 22) = 4.47, p < .05$ ]: remembered items ( $M = -1.06 \mu\text{V}$ ) were more negative than forgotten ones ( $M = -0.71 \mu\text{V}$ ). There were no other interactions, but pairwise comparisons from the three-way interaction showed that, while there were no differences for the initial presentations, remembered spaced repetitions ( $M = -1.07 \mu\text{V}$ ) were more negative than forgotten spaced repetitions ( $M = -0.52 \mu\text{V}$ ) [ $p < .05$ ] and both remembered ( $M = 0.12 \mu\text{V}$ ) and forgotten ( $M = 0.51 \mu\text{V}$ ) massed repetitions [ $p < .01$ ]. No effects of latency were found.

The LPC has been linked to conscious stimulus recognition in repetition paradigms, and may also show subsequent memory effects if the retrieval of information helps long-term memory performance. For voltage, there was a significant interaction of spacing  $\times$  presentation [ $F(1, 22) = 5.38, p < .05$ ]. Massed repetitions ( $M = 4.02 \mu\text{V}$ ) were more positive than all other conditions (spaced repetition:  $M = 3.2 \mu\text{V}$ ; massed initial:  $M = 3.06 \mu\text{V}$  [ $p < .05$ ]; marginal for spaced initial:  $M = 3.27 \mu\text{V}$  [ $p = .055$ ]). Additionally, there was a main effect of memory for voltage, showing a typical subsequent memory effect: recalled ( $M = 3.62 \mu\text{V}$ ) were more positive than forgotten ( $M = 3.16 \mu\text{V}$ ).

Examining the latency of the LPC peak, there was a significant interaction of spacing  $\times$  presentation [ $F(1, 22) = 5.83, p < .05$ ]. Massed repetitions ( $M = 564 \text{ ms}$ ) peaked earlier than all other conditions (spaced repetition:  $M = 597 \text{ ms}$ ; massed initial:  $M = 597 \text{ ms}$ ; spaced initial:  $M = 587 \text{ ms}$  [ $p < .05$ ]).

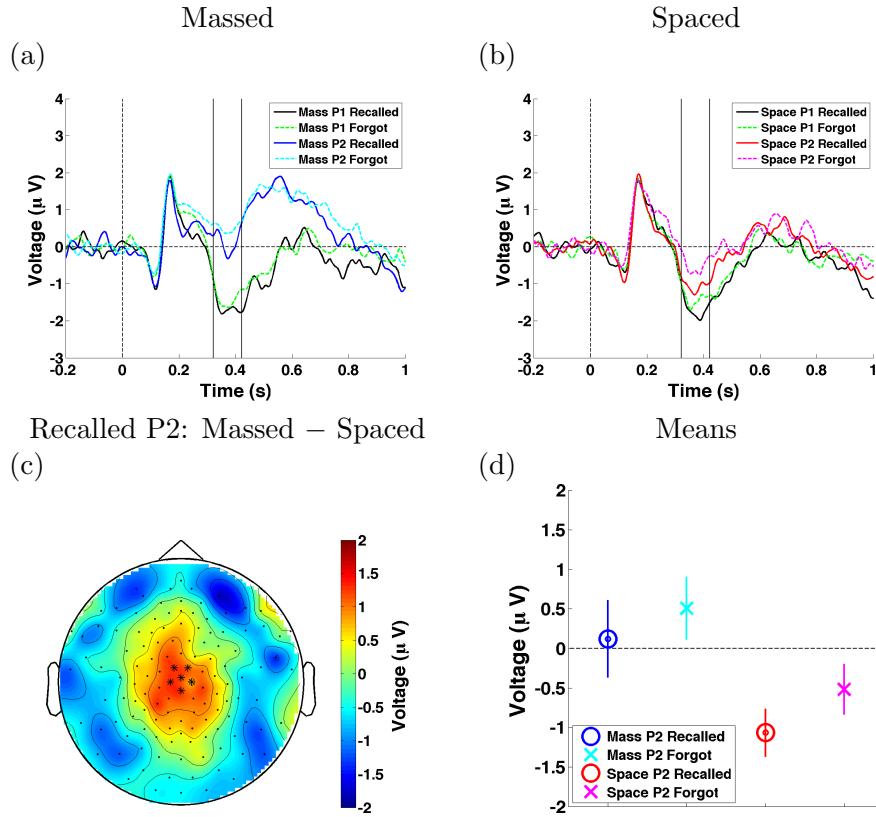


Figure 3.4: N400 to words at electrode Cz and neighbors, analyzed window 322–422 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed *vs.* spaced Presentation 2 (P2) across time and electrodes (\*); (d) analyzed means (error bars are SEM). The negative peak at 400 ms is significantly smaller for massed compared to spaced repetitions.

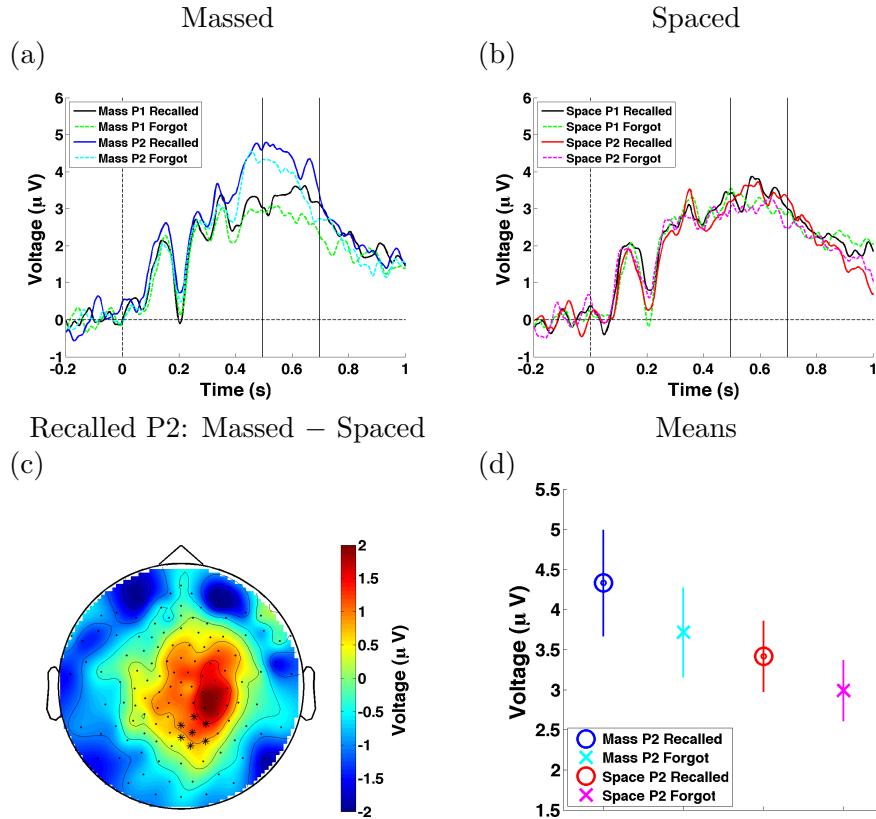


Figure 3.5: LPC to words at electrode 77 and neighbors, analyzed window 496–696 ms: (a) Massed ERPs; (b) spaced ERPs; (c) contrast between subsequently recalled massed *vs.* spaced Presentation 2 (P2) across time and electrodes (\*); (d) analyzed means (error bars are SEM). The positive peak around 600 ms is significantly larger for massed compared to spaced repetitions.

### 3.2.3 ERP discussion

The N1 showed that recalled spaced word repetitions were more negative compared to the other conditions (repetitions of recalled massed words and forgotten spaced and massed words), which can be interpreted to mean that they received the most attention. Massed items are at a disadvantage compared to remembered spaced items, supporting the deficient processing hypothesis. Because remembered massed items are the least negative but are still remembered later, there are likely other mechanisms that influence memory encoding in addition to attention. This implies that attentional processing affects subsequent memory differently for massed and spaced items.

The N400 is more negative when semantic processes are engaged to a greater extent. There was a typical repetition effect showing that voltage decreased with repetition lag, and this interacted with presentation (repetitions attenuate but more so for massed). This implies that semantic processing engages more for spaced than massed repetitions, but not as much as initial presentations. Because the significant effect of subsequent memory showed that remembered items have greater negative amplitudes, this indicates that more semantic processing occurs for remembered items. Spaced items seem to have an overall processing advantage, or perhaps massed items have an overall disadvantage due their attenuated N400. This is all in line with the semantic activation hypothesis of Challis (1993) and the position taken by Van Strien et al. (2007), and therefore supports the deficient processing theory.

The typical interpretation of the LPC effect is that it is more positive during the retrieval of an item's prior presentation (Olichney et al., 2000; Van Strien et al., 2007). The present results show this positivity for massed repetitions as well as an earlier peak, implying faster retrieval of the initial stimulus presentation compared to spaced repetitions. In relation to the N400 results (showing that semantic processing does not need to engage for primed representations), it seems possible that the LPC indexes the information that is in working memory. This result does not seem to directly support any of the theories, unless the match to working memory is an indicator for processing to disengage (deficient processing).

### 3.2.4 Time-frequency results

Spectral decomposition for time-frequency analyses (no low-pass filter) used a set of 38 Morlet wavelets that were logarithmically spaced from 3 to 80 Hz; each wavelet had a width of 6 cycles. Trials were down-sampled to 50 Hz after calculating power and were z-transformed relative to the distribution of all word stimuli with a reference time of  $-300$  to  $-100$  ms relative to stimulus onset. Only the pre-stimulus periods for words were used because the analogous reference time for images would be while words were on the screen. The following frequency bands were analyzed: theta (4.1 to 7.7 Hz); lower alpha (8.4 to 10.1 Hz); upper alpha (11 to 12 Hz); lower beta (13.1 to 20.5 Hz); upper beta (22.4 to 29.2 Hz); lower gamma (31.9 to 45.5 Hz); upper gamma (49.7 to 77.4 Hz).

Time-frequency differences were assessed using repeated measures ANOVAs due to the large number of factors, but because the effect topographies are not as well defined as in ERP analyses a cluster-based permutation test (Maris & Oostenveld, 2007) was used as the basis for electrode choice. Clustering was done by performing a *t*-test for conditions of interest within each time/electrode bin across subjects, followed by grouping together the adjacent bins which yielded a *p*-value of less than .05. Significant differences between pairwise conditions (spaced/massed repetitions, subsequently recalled/forgotten; run separately for words and images) were calculated using a Monte Carlo-style permutation test of the summed *t*-values within a given cluster. Each observed cluster was subject to 500 random permutations of condition labels where its significance was estimated by the proportion of random permutations which yielded clusters that had a summed *t*-value as large or larger than the observed cluster. A given electrode for a given frequency band was included in the analysis if it showed a significant difference in at least two of the six pairwise contrasts. The resulting topographies largely agree with those reported in the memory and attention literature cited in the present paper.

Since oscillatory effects can spread out over time (especially at low frequencies), we analyzed images in addition to words. Because word and image stimuli are presented successively, we need to consider that different patterns may be expected during these to-be-associated stimuli for spaced

and massed repetitions. During a repetition trial, there is the potential for both episodic memory, attentional processes, and semantic processes to occur, perhaps at the same time. Using reverse inference, these can be examined by frequency band (theta, lower alpha, and upper alpha and lower beta, respectively).

Three-way ANOVAs with factors of spacing (spaced and massed), subsequent memory (recalled and forgotten), and time (0–500 ms and 500–1000 ms) were performed for word and image repetitions on power in the theta, lower alpha, upper alpha, and lower beta bands (eight ANOVAs). The two types of stimuli were analyzed separately because their time courses are not necessarily comparable. Only repetition events were analyzed.

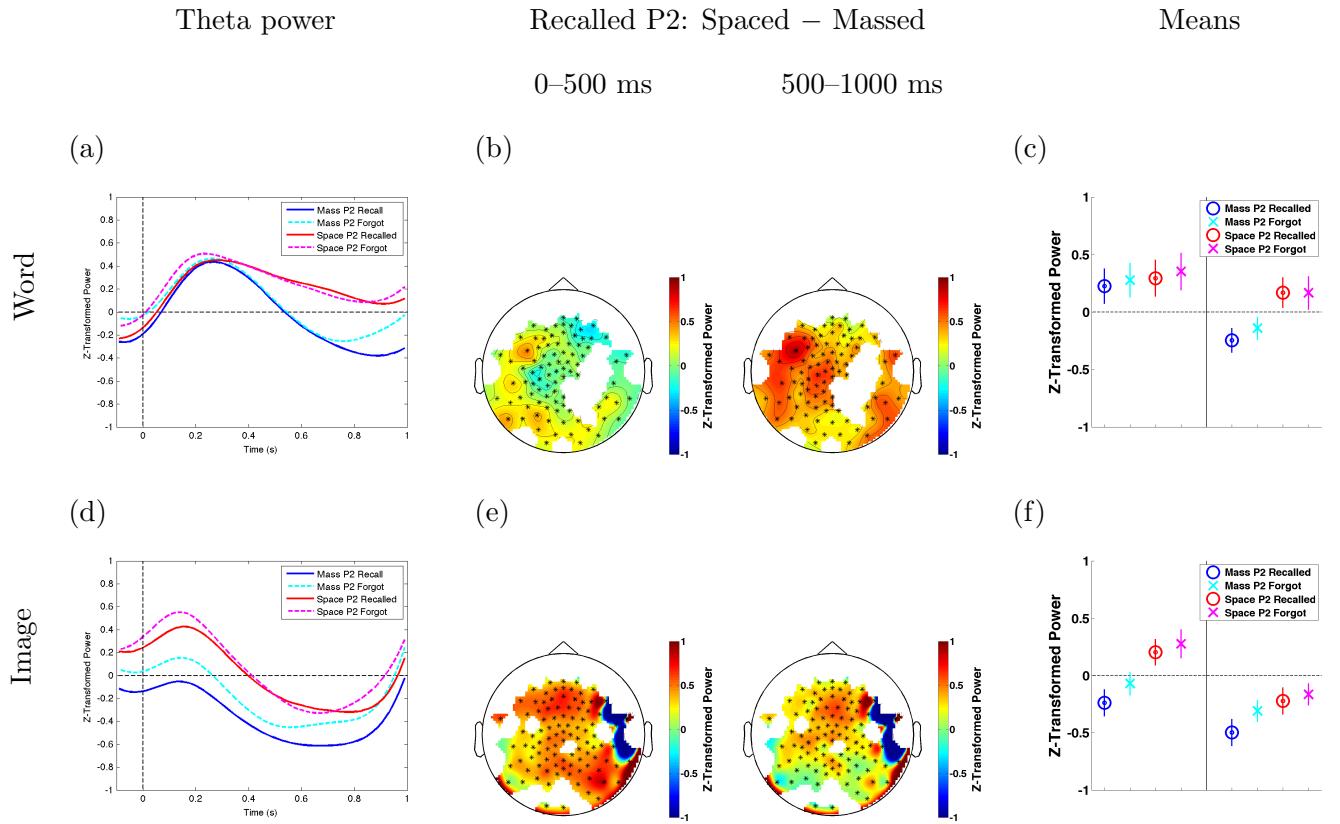


Figure 3.6: Theta power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (\*); (c) and (f) mean values for the two time windows (error bars are SEM).

*Word, theta:* Across 89 electrodes (Figure 3.6, top), theta showed a spacing  $\times$  time interaction [ $F(1, 22) = 26.8, p = 3.43e^{-5}$ ] such that spaced words maintained synchrony across the time windows ( $M = 0.341$  to  $M = 0.163$ ) while massed repetitions showed a power decrease ( $M = 0.267$  to  $M = -0.204$ ); spaced had greater power than massed in the second time window. There were main effects of spacing [ $F(1, 22) = 10.3, p = .00399$ ] and time [ $F(1, 22) = 23.6, p = 7.4e^{-5}$ ] following the same patterns.

(TODO: was there anything for memory spaced gt massed in 3-way pairwise?)

*Image, theta:* Across 99 electrodes (Figure 3.6, bottom), theta continued to show a spacing  $\times$  time interaction [ $F(1, 22) = 6.76, p = .0163$ ] such that spaced images showed greater power than massed in the first time window ( $M = 0.239$  vs.  $M = -0.158$ ). Power dropped more for spaced than massed across the time windows: spaced (second window:  $M = -0.193$ ) showed a difference of 0.432 while massed (second window:  $M = -0.399$ ) showed a difference of 0.241. There were also main effects of spacing [ $F(1, 22) = 31.2, p = 1.28e^{-5}$ ] and time [ $F(1, 22) = 30.6, p = 1.46e^{-5}$ ] following the same patterns, as well as a main effect of memory [ $F(1, 22) = 6.1, p = .0217$ ] showing that less power is associated with subsequent recall ( $M = -0.189$  vs.  $M = -0.067$ ; a negative SME).

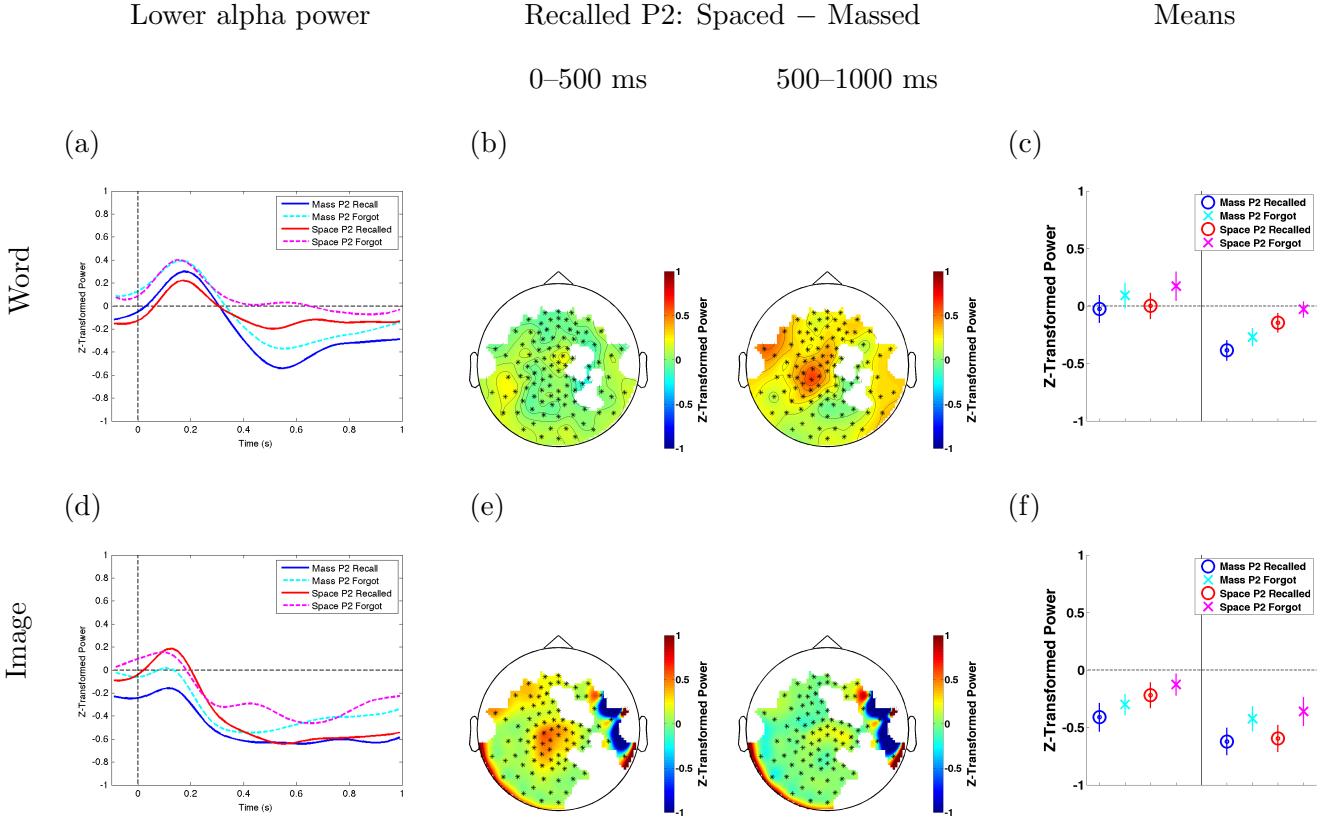


Figure 3.7: Lower alpha power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (\*); (c) and (f) mean values for the two time windows (error bars are SEM).

*Word, lower alpha:* Across 101 electrodes (Figure 3.7, top), lower alpha showed a spacing  $\times$  time interaction [ $F(1, 22) = 8.77, p = .0072$ ] such that massed words showed decreased alpha power compared to spaced in the second time window ( $M = -0.323$  vs.  $M = -0.087$ ). There were also main effects of spacing [ $F(1, 22) = 7.64, p = .0113$ ] (massed showed less power), subsequent memory [ $F(1, 22) = 5.4, p = .0298$ ] (recalled words showed less power), and time [ $F(1, 22) = 15.3, p = .00074$ ] (less power in the later time window).

*Image, lower alpha:* Across 89 electrodes (Figure 3.7, bottom), lower alpha for images showed similar results: a spacing  $\times$  time interaction [ $F(1, 22) = 5.19, p = .0329$ ] such that massed images had decreased power compared to spaced in the first time window ( $M = -0.352$  vs.  $M = -0.183$ ) but not the second ( $M = -0.533$  vs.  $M = -0.497$ ). There were also the same main effects

of spacing [ $F(1, 22) = 4.34, p = .0491$ ] (massed showed less power), subsequent memory [ $F(1, 22) = 12, p = .0022$ ] (recalled words showed less power), and time [ $F(1, 22) = 21.4, p = .000131$ ] (less power in the later time window).

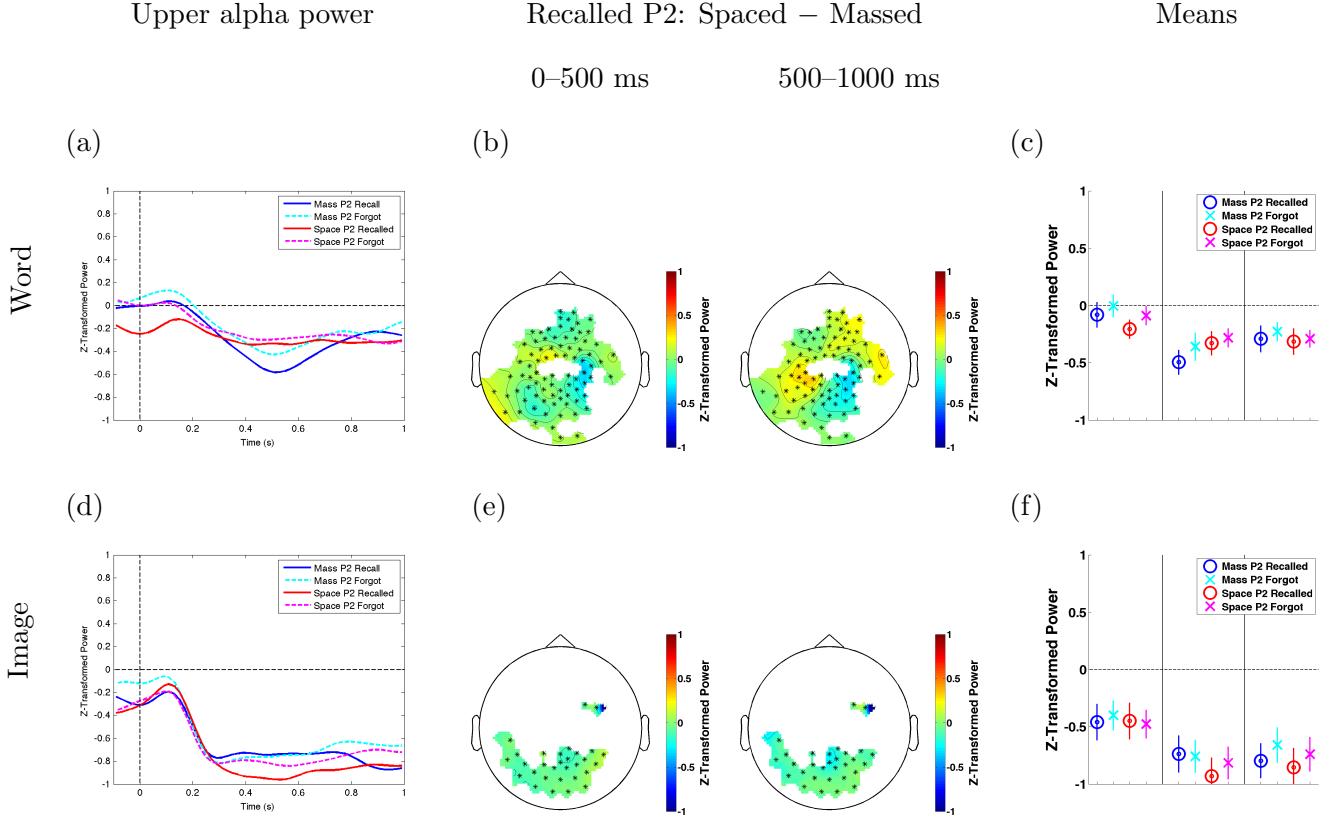


Figure 3.8: Upper alpha power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (\*); (c) and (f) mean values for the three time windows (error bars are SEM).

*Word, upper alpha:* Across 82 electrodes (Figure 3.8, top), upper alpha only showed a main effect of time for words [ $F(1, 22) = 4.47, p = .0461$ ]; there was less power in the second time window than the first ( $M = -0.309$  vs.  $M = -0.183$ ). However, because the data clearly show a negative peak for recalled massed repetitions near 500 ms we divided time into three successive windows. There was a robust spacing  $\times$  time interaction [ $F(2, 44) = 7.46, p = .0038$ ]. Massed were more

negative in the middle time window than in neighboring time windows [ $p < .05$ ], and were more negative than spaced in the middle time windows [ $t(22) = 2.624, p = .015$ ].

*Image, upper alpha:* Across 30 electrodes (Figure 3.8, bottom) and under the three time window ANOVA, upper alpha showed main effects of spacing [ $F(1, 22) = 5.24, p = .032$ ] (spaced had decreased power) and time [ $F(2, 44) = 34.2, p = 3.01e^{-7}$ ] (less power in the two later time windows). There was also a significant three-way interaction [ $F(2, 44) = 3.28, p = .05$ ]: recalled spaced repetitions were more negative in the middle time window than recalled massed repetitions in all time windows [ $p < .005$ ], and recalled spaced repetitions in the late time window were more negative than all massed items in the first two windows [ $p < .05$ ].

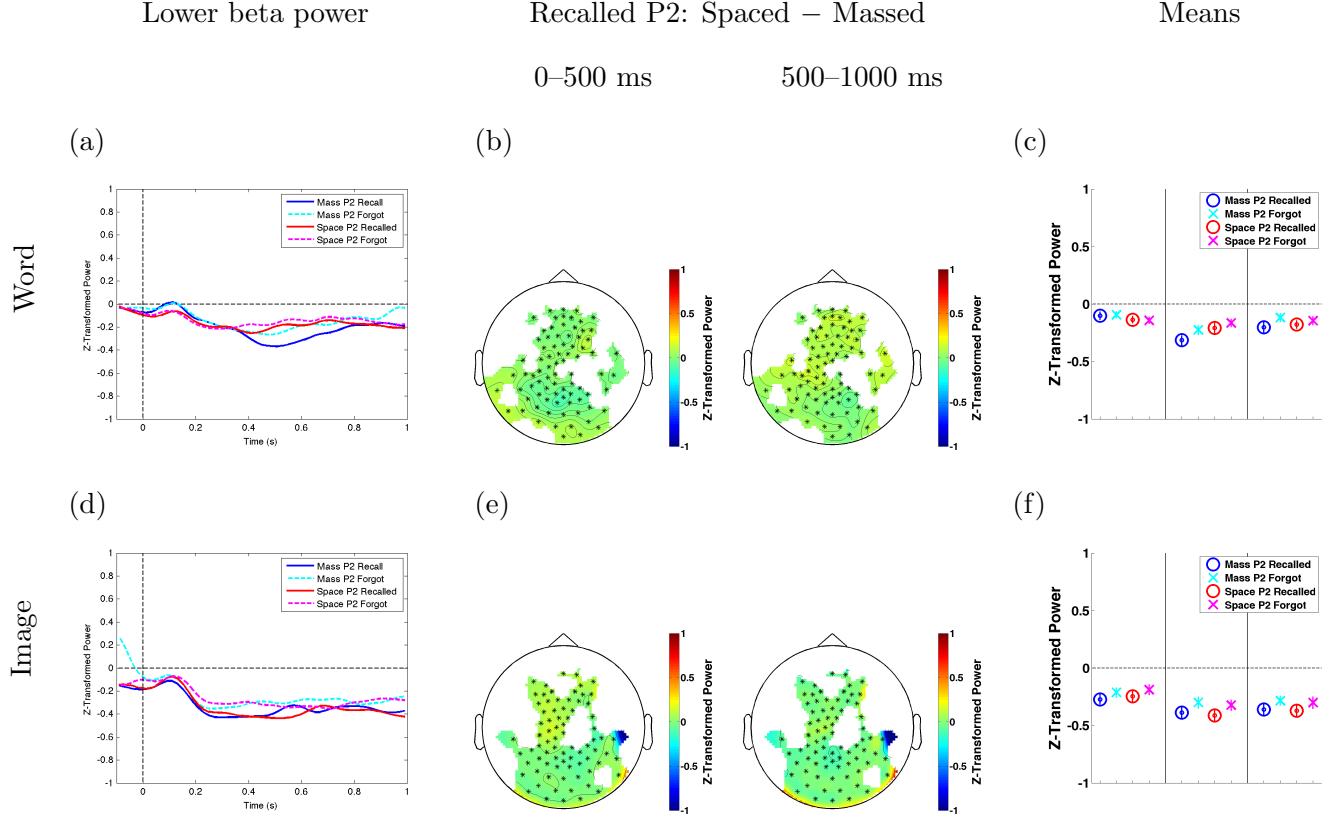


Figure 3.9: Lower beta power to words and images. (a) and (d) Grand averages; (b) and (e) contrast topographic plot between spaced and massed subsequently recalled trials displaying only the analyzed electrodes (\*); (c) and (f) mean values for the three time windows (error bars are SEM).

*Word, lower beta:* Since lower beta oscillations showed a similar dip-and-rise pattern to upper alpha, we also used three time windows in these ANOVAs. Across 81 electrodes (Figure 3.9, top), lower beta showed a spacing  $\times$  time interaction [ $F(2, 44) = 9.89, p = .000453$ ]: massed decreased most in the middle time window [ $p < .01$ ] but spaced showed no differences across time. There was also a marginal memory  $\times$  time interaction [ $F(2, 44) = 2.91, p = .0725$ ]: power for both recalled and forgotten stimuli was lowest in the middle time window [ $p < .05$ ], and was most negative for recalled [ $p < .01$ ]. A main effect of time [ $F(2, 44) = 12.9, p = 5.28e^{-5}$ ] and a marginal main effect of subsequent memory that followed the same patterns [ $F(1, 22) = 4.16, p = .0536$ ].

*Image, lower beta:* Across 78 electrodes (Figure 3.9, bottom), lower beta showed main effects

of memory [ $F(1, 22) = 20.2, p = .00018$ ] (recalled had a power decrease) and time [ $F(2, 44) = 29.3, p = 1.14e^{-8}$ ] (lowest power in the middle time window).

### 3.2.5 Time–frequency discussion

Overall, there was higher theta power for spaced compared to massed repetitions during both words (late) and images (early). Under study-phase retrieval, spaced repetitions will naturally require retrieval from long-term memory, thereby needing to engage processes related to episodic memory. Since massed items are still primed and in working memory we would not expect these processes to engage, and should see theta power drop for these items. The increased theta power for spaced repetitions, particularly through the second time window, denotes that recollection processes are engaged. Contextual variability theory does not require this retrieval, but this and study-phase retrieval would posit that spaced items will be encoded (or re-encoded in the case of study-phase retrieval) with the evolved contextual state, so theta power should be maintained into the image presentation during processes that involve word–image binding (since participants were asked to link the stimuli). Thus, it seems that associative retrieval engages during words and the re-encoding of word–image associations (binding) engages during images; this combination supports study-phase retrieval.

Both positive and negative SMEs have been demonstrated previously in the theta band, so the negative SME for images in theta, which was driven by massed items, is not surprising because it is negative. Rather, it is somewhat surprising because the effect is for massed items. Staudigl and Hanslmayr (2013) changed the direction of the theta effect by varying the context in which words were presented: repetitions in the same context showed a positive SME, while those in a different context showed a negative SME. Though the present results show a negative SME for massed repetitions, it seems unlikely that there are contextual differences for these items because these are immediate repetitions. Perhaps remembered massed stimuli were re-encoded in a more variable fashion (e.g., thought about the stimuli in different ways on each presentation).

The lower alpha band should correlate negatively with general attentional processes and

should be widespread over the scalp. Deficient processing would predict increased alpha power for massed repetitions. We would also expect an effect of memory such that decreased power should be associated with better subsequent memory. Lower alpha showed a sharp decrease during massed words that continued through images compared to spaced (going against the deficient processing prediction), as well as a main effect of memory. This result indicates that more attention gets allocated to massed items and to subsequently recalled items overall. In contrast to the visual N1 ERP component reported earlier, which is also related to attention, lower alpha does not show that massed repetitions are put at an attentional disadvantage; however, the timing difference between these effects implies that they are likely related to different mechanisms. Nonetheless, lower alpha shows a different pattern from our deficient processing predictions. Perhaps these effects follow the LPC and reflect attentional mechanisms accessing prior presentations of massed items more easily than spaced repetitions.

Effects in the upper alpha band and lower beta band should occur late during the word and into the image presentation while semantic processes are engaged (associating the word and the image). Because the semantic representation of massed items is still active, accessing this information may occur faster than for spaced items (like the LPC effect) but may not be processed as deeply. We would also expect an effect of subsequent memory such that trials with more semantic processing (decreased power) are remembered better. However, deficient processing would predict increased power, leading to decreased semantic processing of massed repetitions.

For upper alpha, we found a spacing  $\times$  time interaction after dividing power for words into three time windows; here, massed word repetitions desynchronized earlier and to a greater extent than spaced. This supports the idea that semantic information for massed repetitions is accessed more quickly than for spaced. Perhaps semantic information is quickly sought and retrieved for massed repetitions. Upper alpha for words quickly returns to near-baseline levels (meaning it is inhibited because it no longer needs to be accessed) and spaced repetitions decreased in power (desynchronize) more during the image, supporting a deeper semantic processing of spaced word–image pairs overall. Importantly, when considering three time windows for images the overall

pattern in the three-way interaction showed that recalled spaced image repetitions were the most negative, denoting that increased semantic processing leads to better subsequent memory for spaced trials.

Lower beta showed power decreases for massed words (mostly driven by the recalled trials), and a general subsequent memory effect for images (power decrease is associated with better memory). It seems that increased semantic processing (denoted by decreased power) helps memory overall, and massed trials get a quick semantic processing boost during word presentations but spaced are equally processed during the images. Upper alpha and lower beta power decreases therefore seem to be related to processing semantic information after retrieval, with alpha showing an advantage for spaced and beta correlating more with overall subsequent memory performance.

Across this range of frequency bands, our results show that spaced repetitions involve (a) more retrieval and encoding (theta) starting in the latter half of the word presentation, possibly reflecting the retrieval and encoding of word–image associations, as well as (b) more semantic processing (upper alpha) for remembered spaced repetitions during the image presentation, possibly reflecting the semantic link being made between the word and image.

### **3.2.6 Similarity results**

While the ERP and time–frequency results reveal that massed representations are accessed more quickly but perhaps to a lesser extent semantically than spaced representations, the nature of this representation is still unclear. Are spaced repetitions remembered better due to essentially a repetition effect (in line with study-phase retrieval), or does temporocontextual drift play a role in encoding the repetition in a more variable manner (in line with contextual variability)?

EEG voltages during study image presentations were used to compute the neural similarity between each initial study presentation and its repetition. This same process was done separately for time–frequency oscillations using the following bands: theta, alpha, lower beta, upper beta, lower gamma, upper gamma. Images were used because this is when word–image binding should occur. Neural similarity between study repetitions for individual stimulus presentations was as-

sessed using the method from Manning et al. (2011), described below. All analyzed trials were artifact free, were subsequent hits (correctly recognized as being old), and were divided by the factors of spacing at study (spaced or massed) and subsequent recall at test (correctly recalled or not). The analyzed electrodes were influenced by the regions where Manning et al. (2011) found context-related activity, but as this is a novel analysis the data were manually inspected and regions were chosen by hand (usually over occipital, temporal, and/or parietal regions). Each trial was split into five 200 ms windows (processed and analyzed separately) under the idea that different cognitive mechanisms may occur at different points in time and that these mechanisms may affect the similarity measurement.

Principal Components Analysis (PCA) was used for dimensionality reduction. For each participant, a three-dimensional matrix of voltage measurements (trials by electrodes by time samples) was reshaped into a two-dimensional matrix by unrolling electrodes and time within a trial. Similarly, a matrix with the additional dimension of frequency band was used when analyzing time-frequency data. For each time window, PCA was run on this two-dimensional matrix and the Kaiser criterion was used to choose the components that explained a substantial portion of the variance (eigenvalues  $> 1$ ; Kaiser, 1960). Each retained component is a linear combination of the neural activity (voltage or power for each frequency band) across electrodes and time samples. Each trial then has a weight from each principal component, and together these weights yield the trial's feature vector. The similarity measurement was computed using the normalized dot product of a given item's repetitions (the cosine of the angle between the feature vectors). Finally, the between-trial similarity values for each participant (comparing each event with every other analyzed event, not just its repetition pair) were z-scored so they are in standard deviation units. Thus, a similarity value of zero means the representation is of average similarity compared to all events, and a positive or negative deviation from there means similarity increases or decreases relative to all events.

A three-way repeated measures ANOVA was run on the average voltage similarity values from left and right temporal regions with factors of spacing (spaced and massed), subsequent memory

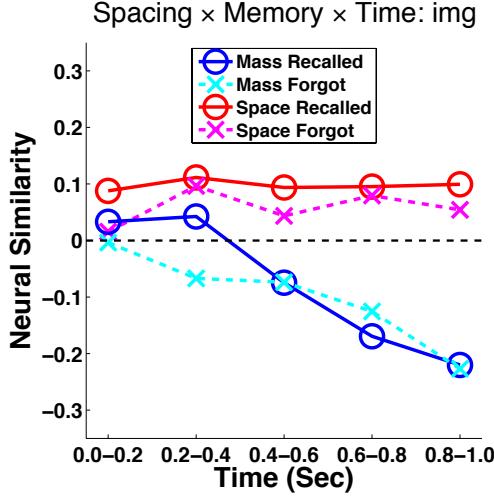


Figure 3.10: Similarity for voltage at left and right temporal sites during image repetitions: Interaction between spacing, memory, and time [n.s.].

(recalled and not recalled), and time (successive time bins). A main effect of spacing was found [ $F(1, 27) = 29.8, p = 8.91e^{-6}$ ] such that spaced presentations ( $M = 0.0777$ ) were more similar than massed ( $M = -0.0886$ ), and a main effect of time was found [ $F(4, 108) = 6.76, p = 8.68e^{-5}$ ] such that neural similarity decreased over time. An interaction between spacing and time explains these effects more fully [ $F(4, 108) = 9.58, p = 2.36e^{-6}$ ]: spaced items kept a relatively consistent level of similarity across the time windows whereas massed items become dramatically dissimilar as time progressed. The three-way interaction was not significant [ $F(4, 108) = 1.18, p = .323$ ] but the data are plotted in Figure 3.10.

The same three-way ANOVA was performed for time-frequency data over left and right parietal regions across all frequency bands. A significant effect of spacing was found [ $F(1, 27) = 7.44, p = .0111$ ], but in the opposite direction of the voltage results: massed ( $M = 0.318$ ) were more similar than spaced ( $M = 0.243$ ). There was also a main effect of time [ $F(4, 108) = 3.42, p = .0116$ ], which was driven by an interaction with subsequent memory [ $F(4, 108) = 6.3, p = .00021$ ]: recalled images showed an increase in similarity at the last time window ( $M = 0.37$ ) compared to forgotten images ( $M = 0.26$ ) [ $t(27) = 2.7614, p = .0102$ ]. Finally, the three-way interaction was marginal [ $F(4, 108) = 2.42, p = .0535$ ] and is plotted in Figure 3.11. The pattern shows that remembered

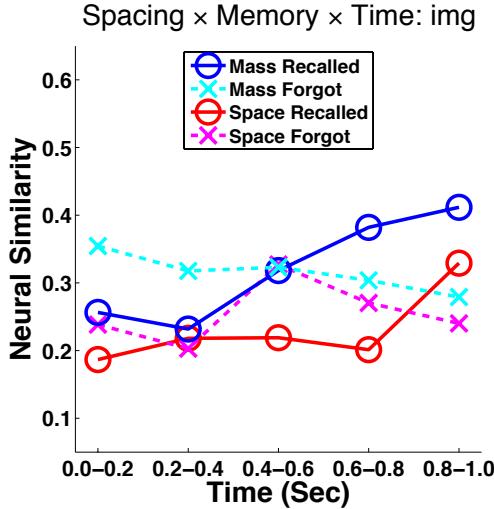


Figure 3.11: Similarity for oscillatory power at left and right parietal sites during image repetitions: Interaction between spacing, memory, and time [ $p = .0535$ ].

items tend to increase in similarity in the second half of the trial while forgotten items tend to decrease; this increase is steeper for spaced compared to massed, though massed seem to receive a similarity boost earlier.

### 3.2.7 Similarity discussion

It is clear that the way massed and spaced stimuli are processed and represented across their repetitions is different. Somewhat surprisingly, the patterns are quite different for voltage and time-frequency data. It is difficult to know exactly what a decrease in similarity means; perhaps it indicates a decrease in processing, or maybe it indicates more variable processing. For voltage, spaced repetitions induced a consistent representation across time, which might support the study-phase retrieval account in that the representation that is being recalled and re-encoded throughout the trial is the same on both presentations. Massed repetitions become much more variable, because either noise is added to the system (a possible explanation supporting deficient processing), they induce a more variable representation (supporting contextual variability), or different neural/cognitive processes are engaged. It seems unlikely that massed representation are more variable because these are immediate repetitions, so contextual drift would not have occurred.

The early attentional ERP results would seem to corroborate support for the deficient processing hypothesis (N1 attenuated for massed), but only if ERPs to words affect processing of the rest of the trial, since similarity during image presentations was analyzed.

For time–frequency data, it seems possible that the greater similarity of massed items speaks to contextual variability, that neural states are more variable at longer delays. Manning et al. (2011), whose methods were used here, calculated the similarity of neural context across study and test and found that it increased with distance, something we might expect to see when measuring the slowly drifting neural state across massed and spaced items. The increase in similarity leading to the subsequent memory effects (three-way interaction) supports study-phase retrieval—or at least the engagement of the same processes—for both massed and spaced repetitions, especially toward the end of the trial. This is in line with another finding by Manning et al. (2011), that context was reinstated when recalling an item during a subsequent test period. Along with recent modeling efforts (Lohnas & Kahana, 2014), these results support the reinstatement of previously studied context (i.e., study-phase retrieval). No other spacing effect accounts seem to be supported with this analysis.

### **3.3 Experiment 1 discussion**

Behaviorally, participants showed a clear spacing effect for both recognition and recall, even using relatively short lags (compared to real-world learning).

Neural activity related to subsequent memory for spaced and massed repetitions is a critical factor to analyze in relation to the spacing effect. The three-way interactions between spacing, repetition, and memory were significant for N1 voltage and upper alpha power, and marginal for time–frequency similarity. Overall, this leads to the idea that increased attention for spaced items (as indexed by the N1) and semantic processing during word–image binding for spaced items (as indexed by upper alpha) benefits subsequent memory.

Early attentional mechanisms orient more to spaced repetitions than massed (N1 was more negative for spaced), especially for those that are subsequently recalled. The N1 may be a signal

to the system for when stimuli should receive additional processing (related to the N400 and upper alpha, discussed below). In fact, using source localization, Proverbio and Adorni (2009) found that the N170 has neural generators in regions that support semantic processing (BA10). This N1 difference supports deficient processing.

A more negative N400 has previously been found to correlate with both semantic processing and subsequent memory, though there is not always a memory effect (reviewed in Friedman & Johnson, 2000). We simply found that voltage was more negative for spaced repetitions compared to massed (and was most negative for initial presentations, though we are most interested in the repetitions). This reflects that semantic processing decreases for massed repetitions, in line with the N1, and also supports deficient processing. The main effect of memory for the N400 shows that subsequent memory is better when semantic processing is engaged to a greater extent during encoding. Matching these results overall, upper alpha showed the important interaction with memory such that remembered spaced repetitions had the largest power decrease, another sign that there is more semantic processing for remembered spaced items.

Our results also support the idea that massed representations are accessed more quickly and to a greater extent on repetition trials than for spaced repetitions: the LPC peaked earlier and was more positive for massed compared to spaced. Here, massed items are easily accessible because they are primed and still in working memory, whereas spaced items need to be retrieved from long-term memory. This latter point is evident in the theta main effect of spacing and in interactions with time; theta is related to memory retrieval and encoding, and spaced words maintained synchrony across time while massed decreased. Because massed repetitions likely feel more familiar due to having greater memory strength, it seems that they are not subjected to additional semantic processing and perhaps are not re-encoded like the theta results suggest spaced items are.

In relation to the neural similarity of massed and spaced presentations (Figure 3.10), massed items decreased in voltage similarity while spaced items maintained a consistent amount. Contextual variability predicts that context has not drifted between massed presentations, but this is not the effect we see. Deficient processing does predict this decrease in similarity. Another possibility

for this pattern of results is the differential activation of neural processes across repetitions (e.g., the same noun is interpreted two different ways for massed items), but combined with the attention (N1) and semantic processing (N400) results, deficient processing seems more likely. For spaced items maintaining similar neural patterns, either the same memory representation is retrieved at the second presentations, the same processes are engaged, or both of these occur. This can support study-phase retrieval, and deeply challenges contextual variability.

The combination of effects for spaced repetitions, from word–image binding (theta oscillation effects) to stable voltage similarity to increased power similarity, seems to support study-phase retrieval for these items that need to be retrieved from long-term memory, and does not yield support for contextual variability. Massed items, on the other hand, tend to be more variable; these EEG effects seem more in line with deficient processing than contextual variability.

## **Chapter 4**

### **Experiment 2**

The goal of Experiment 2 was to replicate and extend Experiment 1 in an attempt to answer some remaining questions about the spacing effect. We examined patterns of effects across short, medium, and long lags to determine whether these spacing effects are modulated by these parameters. For example, deficient processing may still be present at shorter lags (this is important to scrutinize due to its status as an “impostor effect”; Delaney et al., 2010), there may be differences in reinstatement (study-phase retrieval) at longer lags, or we may see other effects change in a graded fashion. Graded effects would qualify more specifically as a lag effect (Greene, 1989; Kahana & Howard, 2005), showing that long-term memory improves as spacing increases. The presence of these gradations would allow us to better interpret data patterns that fit multiple theories. Experiment 1 used lags of 0 and 12; this experiment keeps these lags and adds repetitions at lags of 2 and 32, which are within the range of lags from behavioral spacing studies.

We expected that memory performance would show a lag effect: subsequent memory will correlate positively with lag. The most informative EEG effects regarding the spacing effect in Experiment 1 were for the N1, upper alpha, and time-frequency similarity, though the latter were difficult to interpret. Overall, they implicate differences in attention and semantic processing between spaced and massed repetitions that led to interactions with subsequent memory. It will be important to examine whether these effects modulate with additional lags.

We also included single-presentation stimuli at test, which allowed us to get a baseline measurement of memory performance for comparison of subsequently remembered and forgotten massed

and spaced items. We expected a repetition effect (repeated items will be remembered better than single presentation items), but perhaps if massed repetitions do involve deficient processing then they will be recalled no better than single presentation items.

## 4.1 Method

### 4.1.1 Participants

Forty University of Colorado Boulder undergraduates participated in the experiment for payment of \$15 per hour (ages 18–29,  $M = 21.3$ ; 19 female). All participants were right-handed native-English speakers and had normal or corrected-to-normal vision. Informed consent was obtained from each participant, and the study conformed to the Institutional Review Board guidelines.

### 4.1.2 Materials

The stimuli and experiment presentation software were the same as for Experiment 1.

### 4.1.3 Design

Experiment 2 consisted of two sessions, each of which had nine blocks of three experimental phases: study, distractor, test. Two sessions were used to gather enough trials across the additional conditions. The phases were similar to Experiment 1 (Figure 3.1, page 21). The phase differences were that the exposure phase was excluded to keep the experiment length reasonable and there was no recognition test. The sessions, including application of the electrode net and running in the task, lasted approximately 2.5 hours. Stimuli were randomly shuffled prior to creating the list for each phase at the beginning of the first session. The study phase contained the conditions that were manipulated within subjects, namely the viewing of spaced and massed paired associates.

### 4.1.4 Procedure

An electrode net was applied to each participant’s head, and the first session began with a shortened practice version of the experiment to familiarize participants with the study and test

procedures.

In each study phase block, participants viewed word–image pairs and were asked to think of a relationship between them or to make up a story pairing them. They were told that a subsequent test would require them to remember the word associated with each picture, but they were not told that some pairs will repeat. Spaced items were presented at a lag of either 2, 12, or 32, and massed items were presented at a lag of zero. For each of the two image categories there were three two-presentation spaced pairs per lag, three two-presentation massed pairs, three pairs presented only once, and four additional single-presentation buffers (two at the beginning of the list and two at the end). Buffer pairs did not appear on the test list. Thus, there were 58 word–image presentations in each block. On each trial, a fixation cross appeared for 1.0–1.2 s (jittered), then the word was presented first for 1.0 s followed immediately by the image for 1.0 s. No more than three images from the same category could occur in a row, and no more than two trials with the same lag (including single-presentation pairs) could occur in a row. Each study phase lasted approximately 3.5 min.

In the distractor phase, participants answered simple math problems of the format  $A+B+C=?$  for 2 min. They typed their responses with the keyboard. Different tones occurred for correct and incorrect answers, and mean accuracy and response time was reported to the participant at the end of the phase.

Finally, in the test phase, participants performed a cued recall task. Thirty images (nine spaced, three massed, and three single presentation images from each category) were mixed together and presented one at a time. Participants saw ??????? below each image and had to type the word previously paired with the image; they could pass if they could not remember the word. On each trial, a fixation cross appeared for 1.0–1.2 s (jittered) and the image was shown for 1.0 s, at which point the question marks appeared and participants were asked to make their response. Importantly, test images were presented in a sequence similar to the study order. To construct the test list, the positions of the second presentations of study stimuli were divided into fifteen contiguous groups and each group was shuffled internally. This was done to approximately preserve

a similar amount of time between the second presentation and the test across all “old” stimuli. Each test phase lasted approximately 3 min.

#### **4.1.5 Electrophysiological recordings and data processing**

All procedures for recording and processing electrophysiological data was the same as in Experiment 1, as was finding ERP component peaks and analyzing ERP and time–frequency data.

### **4.2 Results**

Ten participants were excluded from all analyses either because they did not return for the required second session ( $n = 8$ ) or their performance in important conditions was more than 2 standard deviations below the mean ( $n = 2$ ). The remaining thirty participants were included in behavioral analyses. Three additional participants were excluded from ERP and time–frequency analyses either because they had extremely noisy EEG ( $n = 2$ ) or had fewer than 10 artifact-free trials in any of the main trial conditions ( $n = 1$ ), leaving twenty-seven participants in EEG analyses. Similarity analyses included the same participants, all of whom had six or more artifact-free pairs of initial presentation–repetition image trials.

All analyses contingent on subsequent memory are split by whether words were recalled or forgotten. There was no recognition test as in Experiment 1 (during cued recall, all test stimuli were “old”), so the “forgotten” trials category contained both images that were completely forgotten and those where participants were only unsure about the paired words.

#### **4.2.1 Behavioral results**

An ANOVA with factors of session (1 and 2), spacing (single presentation, massed/lag 0, short/lag 2, medium/lag 12, and long/lag 32 spaced), and image category, was run on cued recall hit rate. There was a main effect of spacing [ $F(4, 116) = 174.2, p = 2.62e^{-36}$ ] in the expected order, clearly showing a lag effect: long spaced words ( $M = 56.9\%$ ) were recalled better than medium spaced words ( $M = 49.2\%$ ), and, in turn, performance for short spaced ( $M = 45.8\%$ ), massed

( $M = 35.7\%$ ), and single presentation words ( $M = 23.0\%$ ) was better than the next. Words paired with faces and scenes were recalled at the same rate (faces:  $M = 44.0\%$ ; scenes:  $M = 40.2\%$ ). Thus, there are again clear spacing effects that scale with lag, as well as a simple repetition effect. These rates are comparable to Experiment 1.

There were also session  $\times$  image category [ $F(1, 29) = 7.96, p < .01$ ] and spacing  $\times$  image category [ $F(4, 116) = 7.52, p < .0005$ ] interactions such that recall was better for faces in session 1 compared to session 2, and performance increased at the longest lag for faces compared to scenes. However, these effects do not speak to our investigations of the spacing effect and will not be reported in detail.

Massed items were remembered significantly better than single-presentation items but less well than short (2) spaced items; this still leaves open the possibility that deficient processing occurs for massed items and decreases with lag. Only 9.3 sec elapsed between a short spaced item's initial presentation and repetition, whereas the delay was 3.1 sec for massed items. Neural effects can shed light onto how cognitive processes contribute to the spacing effect. As there were no behavioral effects of session that interacted with spacing, EEG analyses were collapsed over the two sessions.

#### 4.2.2 ERP results

ERP component peaks in Experiment 2 were at the same electrodes as in the previous experiment, but were slightly earlier in time. The visual N1 peaked at electrode 58 (T5) at 144 ms (Figure 4.1;  $\pm 50$  ms window). The N400 peaked at Cz at 352 ms (Figure 4.2;  $\pm 75$  ms window used due a slight spreading of the peak voltage). The LPC peaked at electrode 77 at 536 ms (Figure 4.3;  $\pm 100$  ms window). Again, analyses use these peak electrodes and neighbors during study period words stimuli. After examining condition grand averages, it seems that voltages were overall slightly attenuated compared to Experiment 1; it is hard to give a reason for this, but one possibility is that the electrode cap was placed in slightly different locations for Sessions 1 and 2. Three-way ANOVAs with factors of spacing (massed and short, medium, and long spacing), presentation (1/new and 2/old), and subsequent memory (recalled and forgotten words) were performed on the averaged

window voltages for word presentations. Significant results are reported.

If selective attention is modulated by spacing, early ERP components may show effects; this is relevant to deficient processing. However, there were no significant effects for N1 voltage in the full ANOVA. Because of *a priori* interest based on the results from Experiment 1, we examined the pairwise comparisons in the three-way interaction, which was significant in Experiment 1. Recalled medium (12) spaced repetitions were marginally more negative ( $M = -0.27 \mu\text{V}$ ) than recalled massed repetitions ( $M = 0.03 \mu\text{V}$ ) [ $t(26) = 1.99, p = .057$ ]. The difference compared to massed items was also marginal for recalled short (2) spaced items ( $M = -0.26 \mu\text{V}$ ) [ $t(26) = 2.04, p = .052$ ] but not so for recalled long (32) spaced items ( $M = -0.27 \mu\text{V}$ ) [ $t(26) = -1.26, p = 0.22$ ]. Forgotten conditions were all slightly attenuated. The overall pattern of results is similar to Experiment 1.

If differences in semantic priming and processing contribute to the spacing effect, the N400 should show effects. For N400 voltage there was a significant spacing  $\times$  presentation interaction [ $F(3, 78) = 8.89, p = 6.36e^{-5}$ ] such that voltage became more negative from massed repetitions to spaced repetitions to first presentations [ $p < .01$ ]. There were no lag effects (no quantitative differences between spaced conditions). There were also main effects of spacing (spaced items were more negative than massed [ $F(3, 78) = 11.4, p = 2.85e^{-6}$ ]) and presentation (repetitions were less negative than initial presentations [ $F(1, 26) = 53.2, p = 9.5e^{-8}$ ]). Finally, there was a three-way interaction [ $F(3, 78) = 2.93, p < .05$ ]. There were no differences for the initial presentations. Rather, the effect seems to be driven by recalled medium (12) spaced repetitions showing a negative subsequent memory effect: medium (12) recalled items were more negative than forgotten (recalled:  $M = -0.53 \mu\text{V}$ , forgotten:  $M = -0.19 \mu\text{V}$ ) [ $t(26) = 2.45, p < .05$ ]. The two-way interaction, attenuation for massed compared to spaced trials, attenuation for repetitions, and medium (12) SME are the same patterns as seen in Experiment 1.

The LPC has been linked to the conscious recognition of stimuli in repetition paradigms (particularly those in working memory, as shown in Experiment 1), and may show subsequent memory effects. For LPC voltage, there were main effects of spacing [ $F(3, 78) = 5.68, p < .005$ ], presentation [ $F(1, 26) = 31.6, p = 6.52e^{-6}$ ], and memory [ $F(1, 26) = 22.4, p = 6.88e^{-5}$ ]. Massed

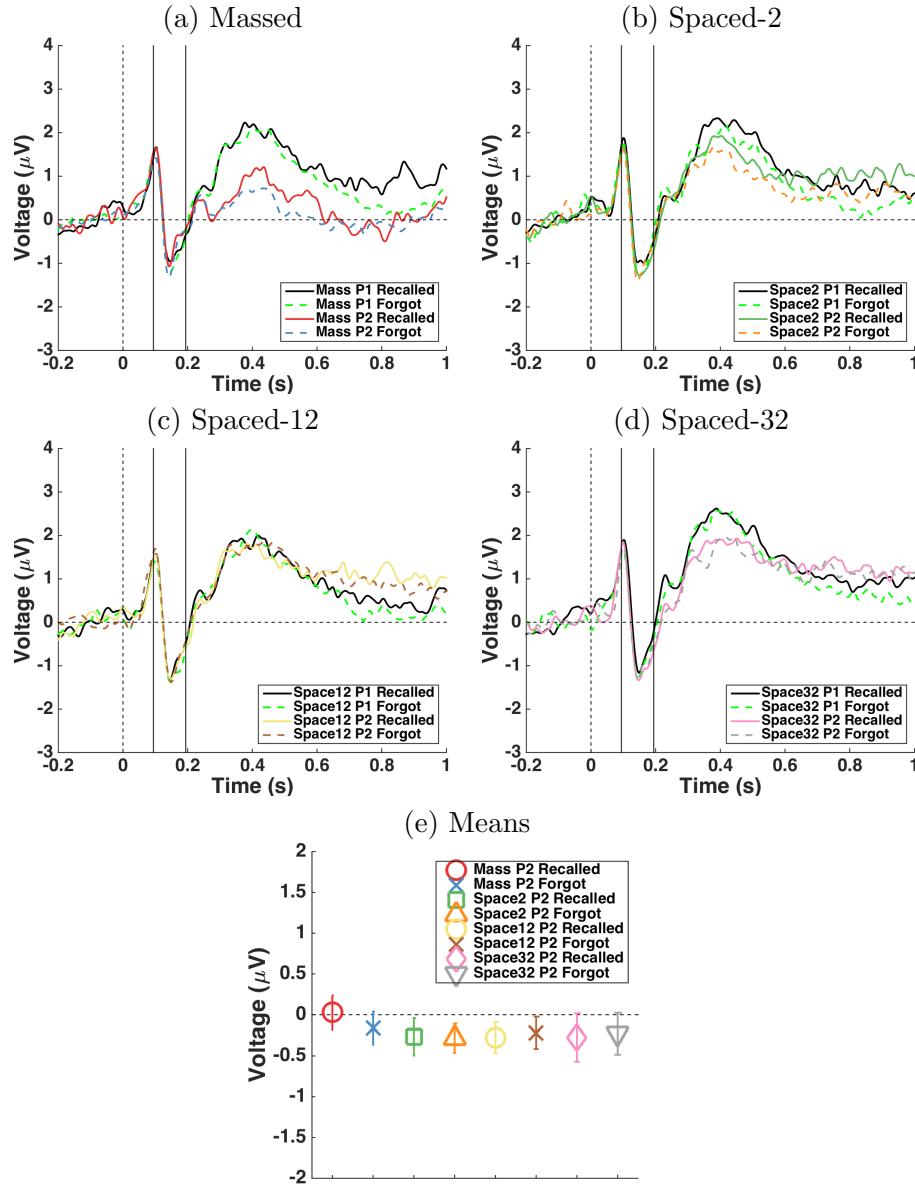


Figure 4.1: N1 to words at electrode 58 (T5) and neighbors, analyzed window 94–194 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The early negative peak is not different across spaced and massed conditions.

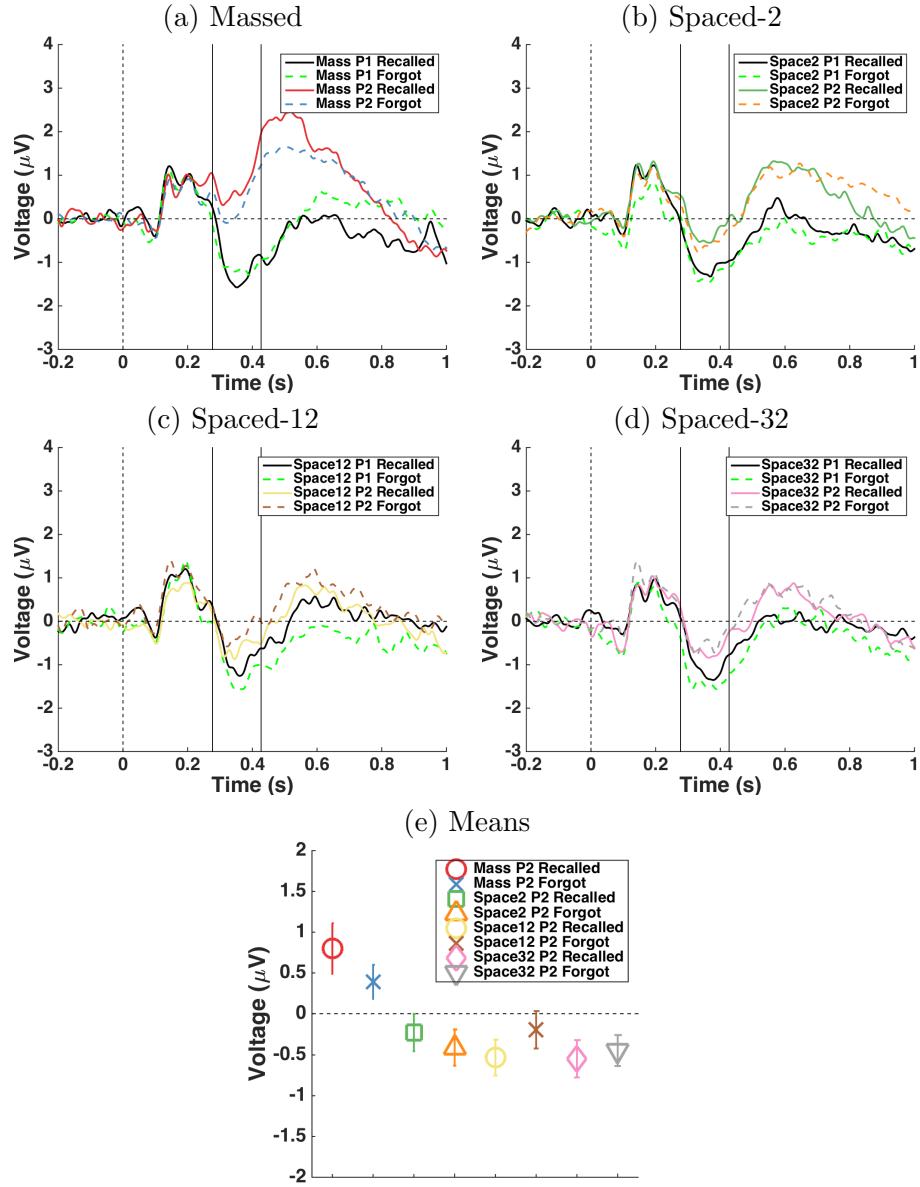


Figure 4.2: N400 to words at electrode Cz and neighbors, analyzed window 227–427 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The negative peak at 400 ms is significantly smaller for massed repetitions compared to any spaced repetition condition.

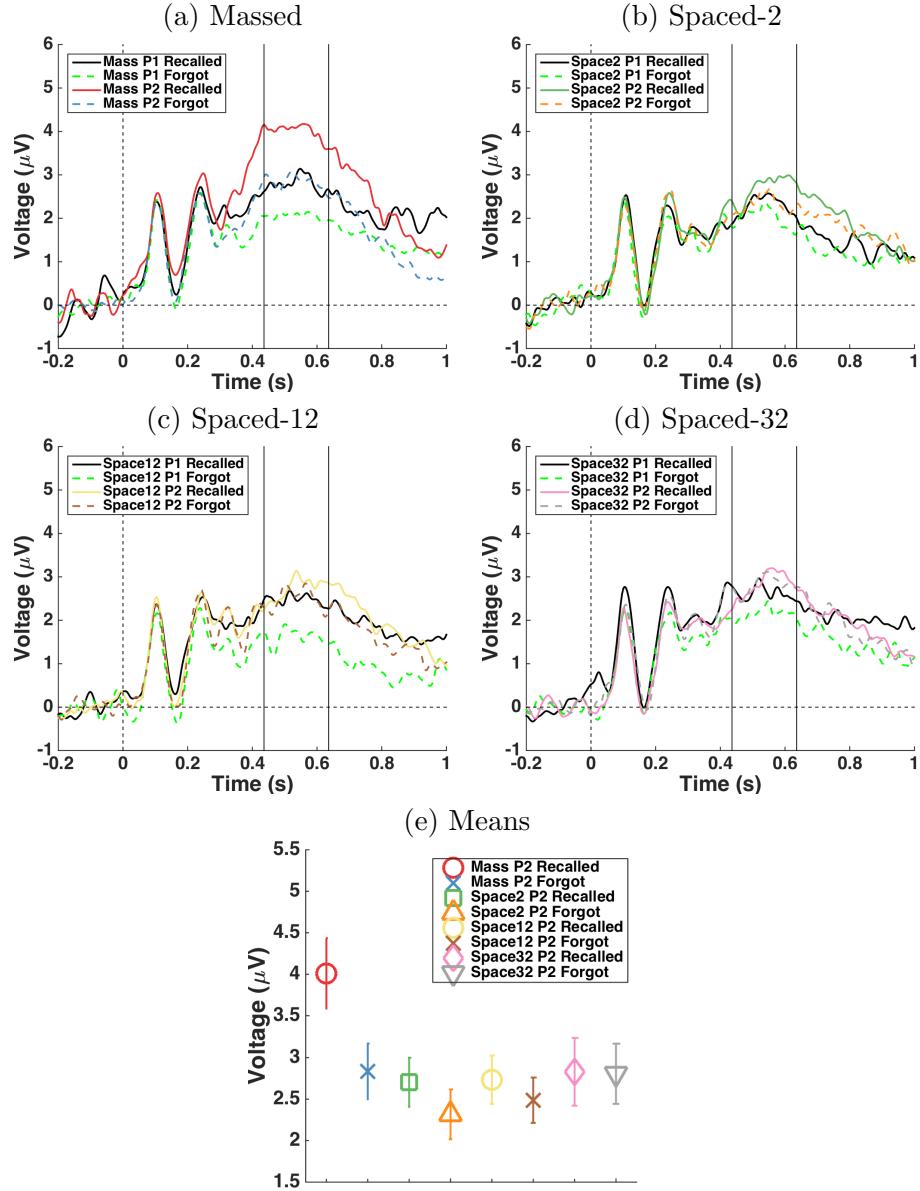


Figure 4.3: LPC to words at electrode 77 and neighbors, analyzed window 436–636 ms: (a) Massed ERPs; (b) short spaced ERPs; (c) medium spaced ERPs; (d) long spaced ERPs; (e) analyzed means (error bars are SEM). The positive peak around 600 ms is significantly larger for massed compared to spaced repetitions.

words ( $M = 2.92 \mu\text{V}$ ) were more positive than short ( $M = 2.33 \mu\text{V}$ ) and medium ( $M = 2.33 \mu\text{V}$ ) spaced words, but only marginally more than long spaced words ( $M = 2.6 \mu\text{V}$ ) [ $t(26) = 1.871, p = .073$ ]; long spaced words were also marginally more positive than short and medium trials [ $p < .1$ ]. On average, repetitions ( $M = 2.4 \mu\text{V}$ ) were more positive than initial presentations ( $M = 2.25 \mu\text{V}$ ), and subsequently recalled words ( $M = 2.81 \mu\text{V}$ ) were more positive than forgotten ( $M = 2.28 \mu\text{V}$ )

. There was no three-way voltage interaction or interaction between spacing and memory in the full ANOVA, but when examining only repetitions using a two-way ANOVA (factors of spacing and memory), there was a significant interaction [ $F(3, 78) = 3.53, p < .05$ ] showing that massed was the only category with a subsequent memory effect; the actual voltage difference between recalled and forgotten words decreased with lag.

For LPC latency effects, which illuminate how quickly information is consciously accessed, there was the same spacing  $\times$  presentation interaction [ $F(3, 78) = 2.86, p < .05$ ] as in Experiment 1: medium (544 ms) and long (550 ms) spaced repetitions peaked later than massed repetitions (529 ms) [ $p < .05$ ]; this pattern was marginal for short spaced repetitions (544 ms) [ $t(26) = 1.9456, p = .063$ ]. Spaced repetitions also peaked later than their respective initial presentations (long: 529 ms [ $t(26) = 3.98, p < .0005$ ], medium: 532 ms [ $t(26) = 1.96, p = 0.061$ ] marginal).

#### 4.2.3 ERP discussion

Under deficient processing we expected the N1 for repetitions to show a lag effect (become more negative as lag increases) because increased early attentional processing should lead to better subsequent memory, for which we have shown a clear behavioral lag effect. For example, a repetition at lag 2 would have a voltage in between a massed item and a repetition at lag 12. However, for N1 voltage there were no significant effects, though examining pairwise comparisons of the three-way interaction revealed qualitatively similar patterns to Experiment 1: spaced items in general seem to get more attentional processing than massed items.

Keeping in mind that the overall pattern from Experiment 1 persisted, the lack of significant N1 effects leads to the idea that attention is not the critical factor for why the spacing effect occurs, at least when defining attention as an early involuntary mechanism indexed by the N1. We cannot make any strong claims in the face of null results, but overall this is less evidence to support deficient processing.

There are a few possibilities to consider for the attenuated N1 effect. Perhaps an experiment difference such as removing the exposure phase made a difference for the N1 in that having an existing representation of a stimulus before needing to learn stimulus pairings could change attentional mechanisms.

Another possibility is that spaced repetitions were relatively more likely to occur in the second experiment (occurred approximately 75% of the time) compared to Experiment 1 (occurred approximately 50% of the time). Perhaps this made spaced repetitions less attention grabbing. We could not find any discussion in the literature relating N1 amplitude to relative stimulus frequency.

A third question regarding the lack of N1 effects is how short term the repetition effects are. Henson, Rylands, Ross, Vuilleumeir, and Rugg (2004) investigated ERP effects for repetitions of pictures of objects at different lags and saw a repetition effect for a similarly timed ERP component (labeled N170, associated with processing faces) after an unfilled 4-second delay (amplitude decreased for repetitions), but not when the four seconds was occupied by another stimulus or at a much longer lag (96 seconds). Thus, deficient processing of a repetition may be eliminated at a relatively short delay if it is filled with other stimuli. Regardless of these differences, based on the results from Experiment 1 we expected an N1 attenuation for massed compared to spaced repetitions, and we still saw strong behavioral spacing and lag effects in Experiment 2, so there seem to be other mechanisms involved in these effects.

We expected the N400 to show lag effects, considering its tie to semantic processing: the component would get more negative as lag increases because more semantic activation is needed during retrieval for longer lags. The N400 showed similar results to Experiment 1, but there were no lag effects across spaced conditions, only a difference between massed and spaced items in

the expected direction: massed items were strongly attenuated compared to spaced. The voltage averages (Figure 4.2e) show an overall pattern of being more negative for remembered items at longer lags. The only significant effect of memory was for recalled medium (12) spaced repetitions, which aligns with Experiment 1, though there is no reason to think this particular spacing lag is important.

Based on the N400, it seems that a similar amount of semantic processing is engaged when there have been at least two intervening stimulus pairs between repetitions. These results imply that spaced repetitions (regardless of lag) receive more semantic processing, or conversely that semantic processing disengages more for massed repetitions. This still supports the Challis (1993) semantic activation hypothesis (less semantic activation for items in working memory), but only one that posits deficient processing for immediate repetitions, and thus cannot explain the behavioral lag effect. Therefore, deficient processing provides only a partial explanation.

Finally, we expected working memory effects (LPC) to be graded across lags. Theoretically, this could occur for two reasons, in opposite directions. First, in terms of memory retrieval (posited under study-phase retrieval), we might see subsequent memory effects for spaced items. More retrieval would be needed as lag increases; it should also be more difficult, as described in the models of Pavlik & Anderson, 2005 and Mozer et al., 2009, but also more beneficial to long-term memory if it is successful. This would posit larger LPC effects as lag increases. Second, the gradation might go in the other direction: effects related to indexing working memory should be stronger for massed items and decrease with lag. This larger difference of subsequent memory for massed items is what we see: massed repetitions were more positive and peaked earlier than spaced repetitions, and showed an SME.

Also supporting the idea of the LPC indexing conscious access to these representations are the latency effects: their later peaks show that it takes longer to access medium and long spaced representations compared to the massed and short spaced conditions. As in Experiment 1, it seems that the LPC indexes the information that is in working memory, and again, the result does not seem to directly support any of the theories unless the match to working memory for massed items

is an indicator of deficient processing.

Though we have found some support for deficient processing, specifically in relation to massed items, the mechanisms involved in this hypothesis seem to have little bearing on the spaced conditions and therefore cannot capture the lag effect. Thus, we still have not found a defining neural signature for why performance increases with lag.

#### 4.2.4 Time–frequency results

As in Experiment 1, we analyzed images in addition to words for time–frequency data. Three-way ANOVAs with factors of spacing (spaced and massed), subsequent memory (recalled and forgotten), and time (0–500 ms and 500–1000 ms) were performed for word and image repetitions on power in the theta, lower alpha, upper alpha, and lower beta bands (eight ANOVAs).

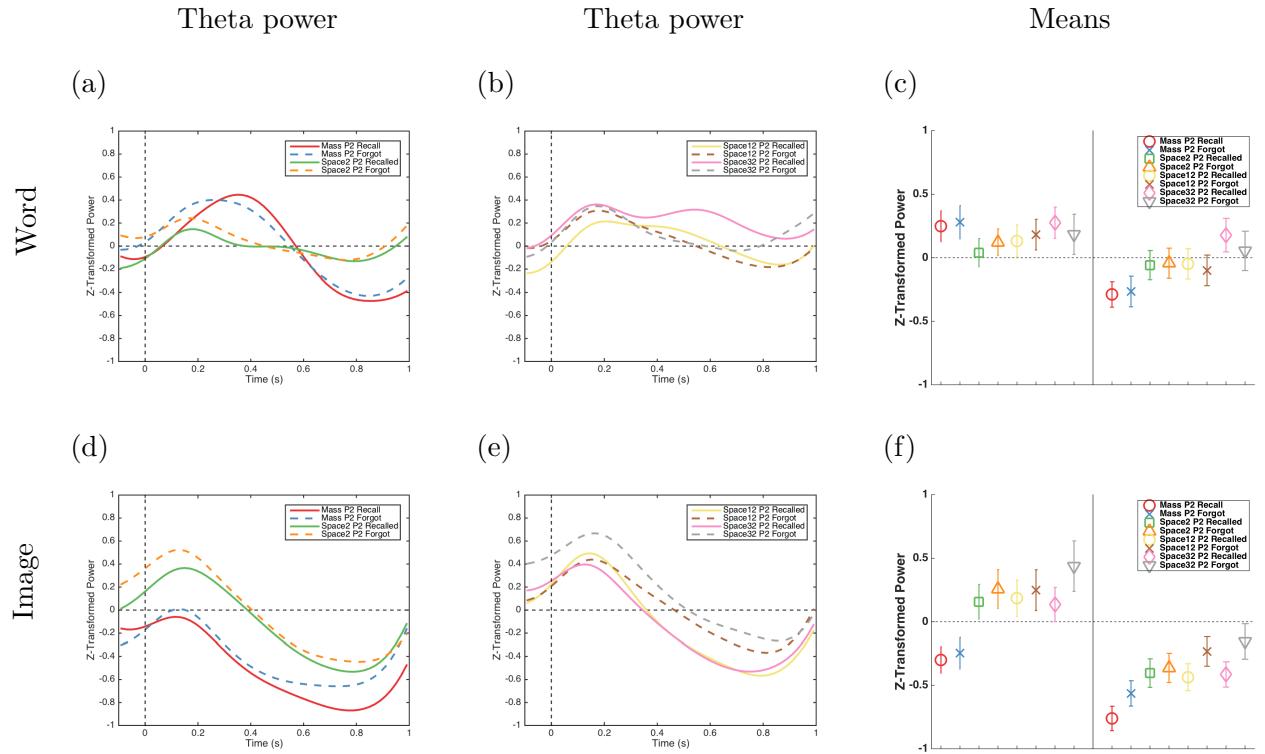


Figure 4.4: Theta power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the two time windows (error bars are SEM).

*Word, theta:* Across 73 electrodes (Figure 4.4, top), there was a spacing  $\times$  time interaction [ $F(3, 78) = 16, p = 5.33e^{-8}$ ] such that all conditions decreased across time [ $p < .05$ ] except long (32) spaced words. Contributing to this were main effects of spacing [ $F(3, 78) = 4.0, p < .05$ ] and time [ $F(1, 26) = 33, p = 4.81e^{-6}$ ]. Long (32) spaced word repetitions showed more theta power than all other conditions [ $p < .05$ ], and there was a decrease in power across time. There was no three-way interaction, but pairwise comparisons for recalled words between massed and spaced conditions in the later time window showed that long (32) spaced words had significantly higher theta power than all other conditions.

*Image, theta:* Across 90 electrodes (Figure 4.4, bottom), there were main effects of spacing [ $F(3, 78) = 27.2, p = 1.17e^{-10}$ ], memory [ $F(1, 26) = 8.76, p < .01$ ], and time [ $F(1, 26) = 91.1, p = 5.57e^{-10}$ ]. Massed items desynchronized more overall than the other conditions, subsequently recalled items desynchronized more than forgotten ones, and power decreased across time. A spacing  $\times$  time interaction [ $F(3, 78) = 5.14, p < .01$ ] showed that the decrease in power across time for massed items was less than short (2) and medium (12) spaced repetitions.

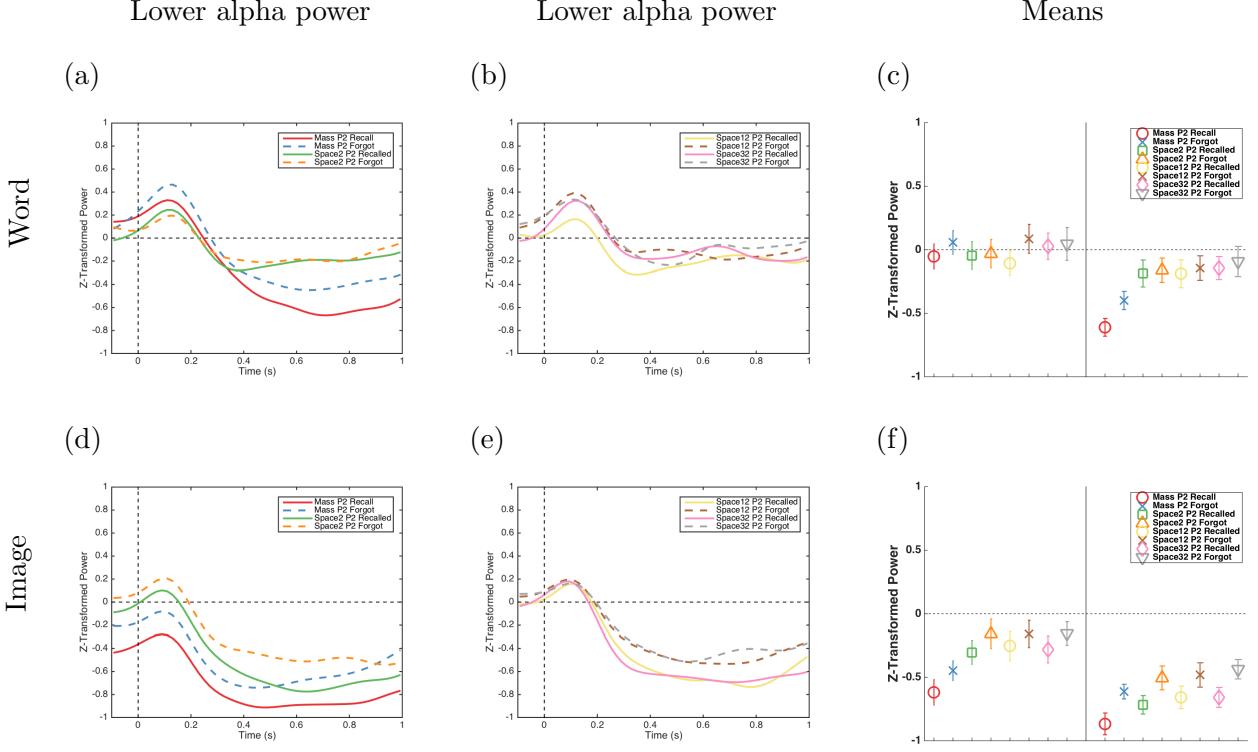


Figure 4.5: Lower alpha power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the two time windows (error bars are SEM).

*Word, lower alpha:* Across 49 electrodes (Figure 4.5, top), we saw the same spacing  $\times$  time interaction [ $F(3, 78) = 15.1, p = 2.15e^{-7}$ ] as in Experiment 1. Lower alpha desynchronized more (power was more negative) for massed compared to spaced words at all lags in the second time window. There were also significant main effects of spacing [ $F(3, 78) = 9.78, p = 2.22e^{-5}$ ], memory [ $F(1, 26) = 5.89, p < .05$ ], and time [ $F(1, 26) = 9.71, p < .005$ ]. Massed word repetitions showed more lower alpha desynchronization than the other conditions, there was more desynchronization for recalled words, and there was more desynchronization in the second time window; these effects were the same as in Experiment 1 and the decreases seem to be driven by massed recalled items.

*Image, lower alpha:* Across 94 electrodes (Figure 4.5, bottom), lower alpha for images showed the same pattern of effects as for words. Massed image repetitions desynchronized more overall compared to spaced repetitions of all lags (main effect of spacing, [ $F(3, 78) = 14.3, p = 2.65e^{-7}$ ]) but the spacing  $\times$  time interaction [ $F(3, 78) = 6.23, p < .005$ ] showed a larger decrease from the

first to the second time window for all spaced conditions compared to massed. There were also main effects of memory [ $F(1, 26) = 24.7, p = 3.67e^{-5}$ ] (remembered desynchronized more than forgotten) and time [ $F(1, 26) = 33.6, p = 4.19e^{-6}$ ] (more desynchronization [lower power] in the second time window). Additionally, there was a memory  $\times$  time interaction [ $F(1, 26) = 5.37, p < .05$ ]; there were SMEs in both time windows, but the difference was larger in the later time window due to recalled items desynchronizing more.

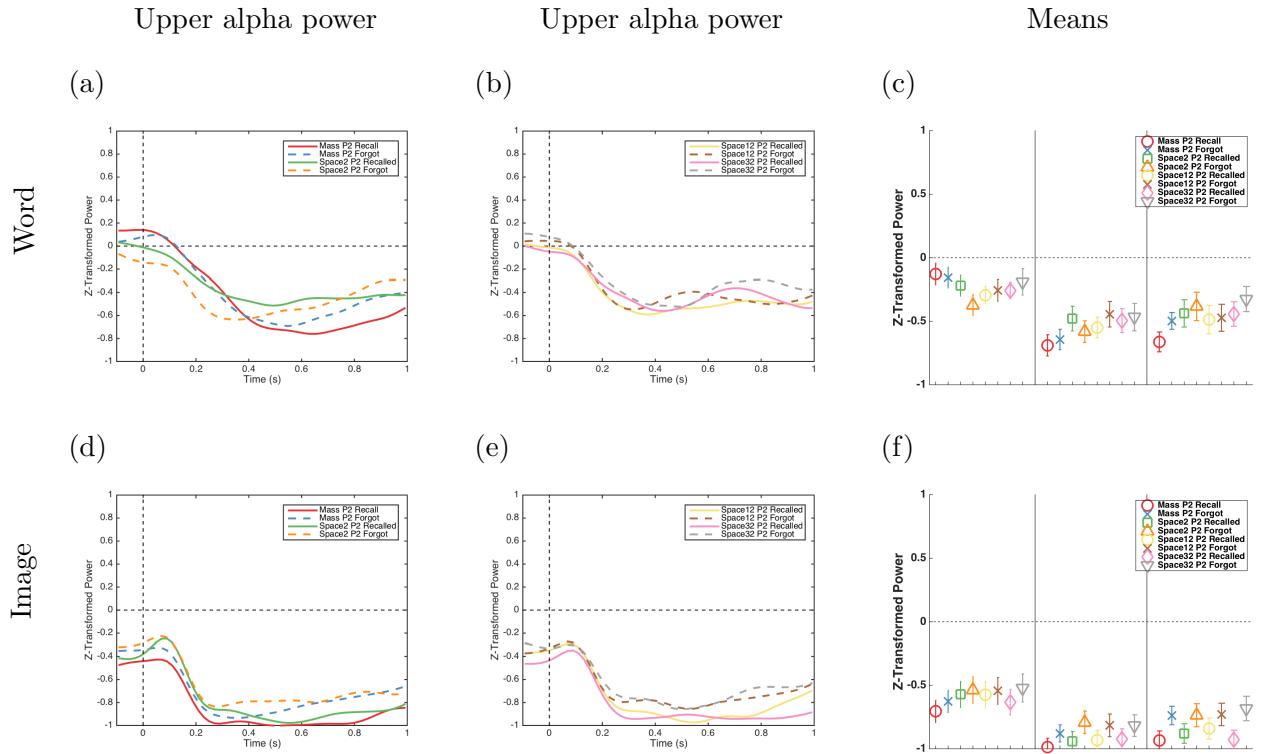


Figure 4.6: Upper alpha to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the three time windows (error bars are SEM).

*Word, upper alpha:* As in Experiment 1, we used three time windows for upper alpha and lower beta. Across 46 electrodes (Figure 4.6, top), there was a main effect of time [ $F(2, 52) = 18.3, p = 2.36e^{-5}$ ]: the second and third time windows showed more upper alpha desynchronization than the first [ $p < .005$ ]. There was also a spacing  $\times$  time interaction [ $F(6, 156) = 11.8, p = 9.2e^{-8}$ ]: each condition decreased from the first to the second time window (and stayed the same for the

third) [ $p < .05$ ], but massed decreased more. Additionally, massed was more desynchronized than all spaced lags for the second and third time window (except medium (12) spaced in the third time window) [ $p < .01$ ].

*Image, upper alpha:* Across 44 electrodes (Figure 4.6, bottom), there were main effects of memory [ $F(1, 26) = 21.8, p = 8.08e^{-5}$ ] (recalled items desynchronized more than forgotten items) and time [ $F(2, 52) = 25, p = 3.55e^{-6}$ ]: power was lower in the middle time window compared to the others [ $p < .05$ ]. The spacing effect was marginally significant [ $F(3, 78) = 2.74, p = .053$ ]: massed images desynchronized more than short and medium spaced items [ $p < .05$ ]. There was also a memory  $\times$  time interaction [ $F(2, 52) = 6.82, p < .005$ ] showing significant negative SMEs in all time windows [ $p < .05$ ] but that the SME was strongest in the late window.

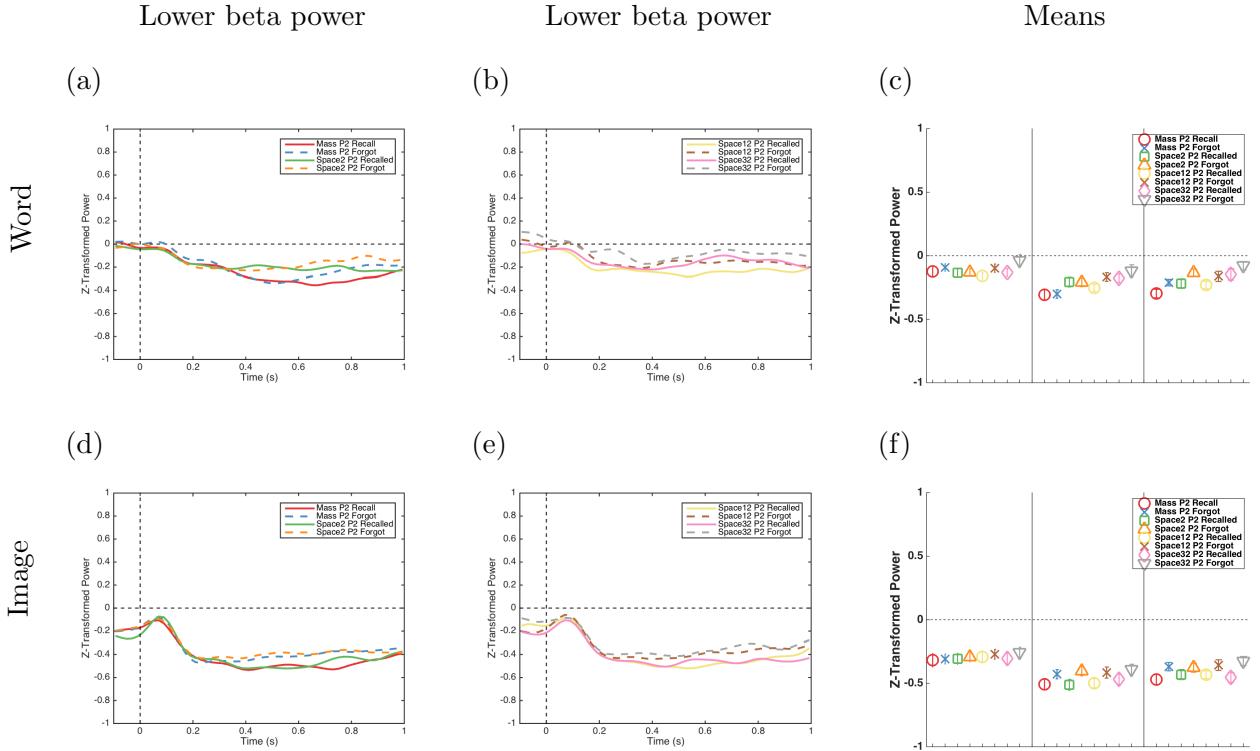


Figure 4.7: Lower beta power to words and images. (a), (b), (d), and (e) Grand averages; (c) and (f) mean values for the three time windows (error bars are SEM).

*Word, lower beta:* Across 66 electrodes (Figure 4.7, top), there was a spacing  $\times$  time inter-

action [ $F(6, 156) = 6.96, p = 3.79e^{-6}$ ]. Spacing effects across all spaced lags (short, medium, and long) were larger in the second and third time windows (massed were more negative). There were main effects of spacing [ $F(3, 78) = 9.48, p = 7.16e^{-5}$ ] and time [ $F(2, 52) = 25.5, p = 5.06e^{-8}$ ] that followed the patterns of the interaction, as well as one of memory [ $F(1, 26) = 8.36, p < .01$ ] Massed word repetitions desynchronized more than the spaced conditions, there was more desynchronization during the second time window compared to the others (the third was lower than the first) [ $p < .05$ ], and subsequently recalled words desynchronized more than forgotten ones.

*Image, lower beta:* Across 30 electrodes (Figure 4.7, bottom), there were main effects of memory [ $F(1, 26) = 25.2, p = 3.22e^{-5}$ ] and time [ $F(2, 52) = 53.7, p = 3.33e^{-12}$ ] which followed the patterns of effects reported above for words: recalled images showed more desynchronization than those that were forgotten, and power decreased into the second time window compared to the first and increased in the third [ $p < .005$ ].

#### 4.2.5 Time–frequency discussion

The theta band is a place where we expect to see effects of recollection and encoding, especially at longer lags where retrieval would be more difficult but also more beneficial to long-term memory if successful (Delaney et al., 2010; Pavlik & Anderson, 2005)). Overall, the theta effects are quite similar to Experiment 1, and there is some variety across lags that is of interest for lag effects. Theta power was sustained across word repetitions for spaced items compared to massed items, particularly for long (32) spaced words. Recalled long (32) spaced words also showed higher power than all other conditions in the second time window. This sustained theta spacing effect is interesting because these trials showed the highest behavioral memory performance. In fact, the pattern of theta power across lags for recalled word repetitions in the second time window follows behavioral performance. Thus, it seems that theta effects support study-phase retrieval in much the same was as Experiment 1, and theta may be an indicator of a lag effect.

For images, there was again a negative SME such that less theta power was associated with recalling the associated words during the test phase. The discussion of this effect in Experiment 1

brought up the idea that it may be due to a more varied contextual state, though this again does not seem to mesh with the idea of little contextual drift occurring for massed repetitions.

The lower alpha band generally correlates negatively with attentional processes (decreased power, increased attention) and shows a widespread scalp topography. Deficient processing would predict increased alpha power for massed repetitions because less attention would be given to these items. However, because the results for Experiment 1 were in the opposite direction of this prediction (massed desynchronized more, or showed a decrease in power) we expected the present results to be the same, and they were. We also expected to see more desynchronization (decreased power) for better subsequent memory, due to increased attentional processing; this was confirmed. The negative SME (subsequently remembered items desynchronized more than forgotten items) indicates that these items received more attention. Finally, the spacing effect again does not support deficient processing: we saw decreased power for massed items, whereas deficient processing would expect less attention to be given to massed items.

Spaced images had a larger power decrease from the first to the second time window compared to massed images, though massed were still lower overall.

For upper alpha, which is associated with semantic processing, we expected to see lag effects: subsequently remembered items should desynchronize more as lag increases because items with longer lag are remembered better on average, and semantic processing should help memory. The results showed negative SMEs in all time windows for image presentations, meaning that remembered items desynchronized more and therefore incur more semantic processing, but there were no lag effects and the spacing effect is opposite from what we predicted. Massed words showed more desynchronization than all spaced lags—implying more semantic processing—starting in the second time window and continuing into the third, so perhaps upper alpha desynchronization is driven by the semantic representation being active in working memory. A potentially important difference from Experiment 1 is that there were no effects showing that spaced items desynchronized more than massed items (this occurred for image repetitions in Experiment 1). (TODO: This is important because...) As in Experiment 1, these results do not support deficient processing because this

theory would predict less semantic processing for massed items.

If upper alpha desynchronization is linked to the retrieval of semantic information, it is not clear why massed items would show a greater power decrease since these items should already be in short term memory. However, both massed and spaced items showed this qualitative pattern, so both likely involve semantic processing.

The lower beta effects for words and images mostly follow those of Experiment 1. The lack of a memory  $\times$  time interaction for words was the main difference, but this is not a critical result. Massed words desynchronized more than spaced words, and subsequently recalled stimuli showed more desynchronization. These results solidify the idea that semantic processing (denoted by decreased power) helps memory overall. Massed trials get a semantic processing boost during word presentations (more desynchronization in the middle and late time windows), perhaps because they are in working memory and the information is readily accessible, but spaced and massed items follow the same pattern during the images, implying that they are processed equally.

#### 4.2.6 Similarity results

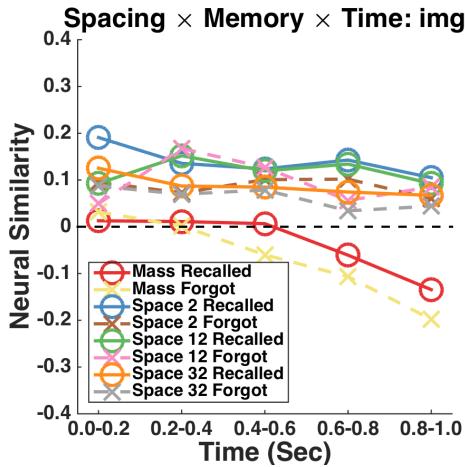


Figure 4.8: Similarity for voltage at left and right temporal sites during image repetitions: Interaction between spacing, memory, and time [n.s.].

None of our effects for Experiment 2 have supported contextual variability; to do so, we would need to see more variability in EEG at as lag increases, an effect of temporocontextual drift. Perhaps a clearer picture will emerge with more lags.

The same ROIs, latencies, and data processing methods were used to measure neural similarity between initial and repetition trials as in Experiment 1, including using images for analysis because this is when word–image (item–context) binding should occur. Again, these analyses measure the neural similarity between the initial presentation and repetition of individual stimuli. All artifact-free spaced and massed trials were included in analyses.

For voltage (Figure 4.8), a three-way repeated measures ANOVA was run on the average similarity values from left and right temporal regions with factors of spacing (spaced and massed), subsequent memory (recalled and not recalled), and time (successive 200 ms time bins). There was a main effect of spacing [ $F(3, 78) = 11.1, p = 5.19e^{-6}$ ], a marginal effect of memory [ $F(1, 26) = 3.92, p = .058$ ], and a main effect of time [ $F(4, 104) = 6.31, p < .0005$ ]. Spaced items of all length lags were more similar than massed items, subsequently recalled items were marginally more similar than forgotten ones, and similarity decreased across time. There was also a spacing  $\times$  time interaction [ $F(12, 312) = 2.99, p = .005$ ]. The interaction explains the two significant main effects in that massed item similarity decreased across time whereas spaced items stayed consistent. These results exactly follow the results of Experiment 1, with the addition of the marginal SME.

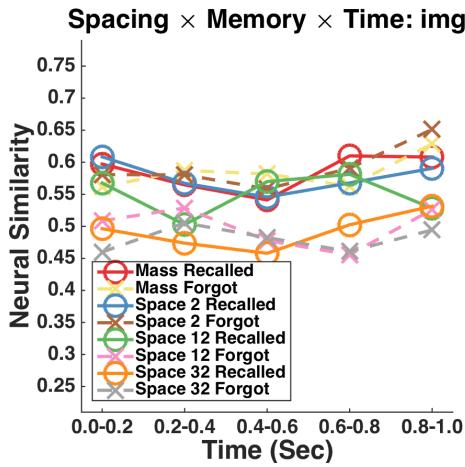


Figure 4.9: Similarity for oscillatory power at left and right parietal sites during image repetitions: Interaction between spacing, memory, and time [n.s.].

For time–frequency data (Figure 4.9), the same three-way ANOVA was performed over left and right parietal regions across all frequency bands. There was a main effect of spacing [ $F(3, 78) = 8.92, p = 6.5e^{-5}$ ] showing that similarity decreased with lag (massed and short (2) spaced repetitions were more similar than medium (12) and long (32) spaced repetitions, [ $p < .05$ ]). There was also main effect of time [ $F(4, 104) = 2.9, p < .05$ ] showing that similarity increased across time. Finally, a marginal memory  $\times$  time interaction [ $F(4, 104) = 2.46, p = .051$ ] showed that recalled and forgotten both increased over time but with different trajectories (overall the interaction pattern is unclear). These results follow Experiment 1.

#### 4.2.7 Similarity discussion

The patterns of neural similarity results for both voltage and time–frequency data closely followed Experiment 1. For voltage, spaced repetitions induced a consistent representation across time, which supports study-phase retrieval under the idea that the same representation is engaged in each presentation. Massed repetitions become more variable, which again might be attributed to not processing the stimuli and effectively adding noise to the system (supporting deficient processing), inducing a more variable representation (supporting contextual variability), or engaging different

neural/cognitive processes. Perhaps the decreased similarity comes from extraneous information being activated during massed presentations, thereby making them more variable. This could align with the stimuli being thought about in a different way, which was proposed regarding the theta band negative SME (main effect). Regarding contextual variability, from a theoretical standpoint there is no good reason for massed repetitions to have more variable representations because context has not drifted. Because we know that, on average, memory for spaced items is better than for massed items, and because consistent representation should occur under study-phase retrieval, this theory seems to help explain why the spacing effect occurs.

For the similarity of time–frequency data, in the same way it was discussed for Experiment 1, it makes sense that longer lags would have lower similarity because context has drifted for these repetitions more than for short lags. This is the cleanest pattern we can make out in a complex set of results—that medium and long spaced items were less similar than massed and short spaced items—a pattern we would expect to see under contextual variability. There was no three-way interaction between spacing, time, and memory as in Experiment 1. The remaining effects seem to neither support nor challenge the hypotheses.

### **4.3      Experiment 2 discussion**

The purpose of this experiment was to look for lag effects in relation to neural activity to better interpret data patterns in the context of the three theories. Behaviorally, we saw exactly what was expected: performance increased significantly as lag increased. Our hope is to discover some of the neural correlates behind these patterns.

Again, we are interested in the neural activity related to subsequent memory for spaced and massed repetitions, as this is the critical point where spaced and massed items differ. Differences between the experiments are covered in the General Discussion section below, but overall the EEG effects followed the patterns of Experiment 1.

Indicators of early attentional processes again revealed that spaced repetitions receive more attention: the N1 ERP component was more negative compared to massed for the medium (12)

spaced condition, which had the same spacing as in Experiment 1. This supports deficient processing of massed repetitions relative to spaced. The N1 may be an indicator for semantic processing to engage, reflected by the N400 and upper alpha band. N400 voltage was more negative for spaced compared to massed repetitions overall. Additionally there was an interaction with memory for the medium (12) spaced condition: recalled repetitions were more negative than forgotten. Thus it seems that memory is better when semantic processing engaged to a greater extent, and this occurs more for spaced repetitions than massed. Upper alpha did not show the three-way interaction (spacing × presentation × memory), but we did see that recalled image repetitions desynchronized more than forgotten images, indicating more semantic processing for the recalled condition.

It seems that the LPC is an indicator of when information is primed and in working memory. Massed repetitions peaked earlier and to a greater extent than spaced repetitions. Interestingly, this varied with lag, though not in a full gradient: the massed peak was only earlier compared to medium and long spaced repetitions, indicating that short spaced items may still be somewhat primed. Also, the spacing × memory interaction showed that the SME decreased with lag, which might indicate less information being recognized at longer lags. Theta aligns with these effects, showing retrieval processes occurring for spaced items (maintaining synchrony across time, specifically for the long (32) spaced repetitions) but not for massed items (desynchronizing across time). This supports study-phase retrieval. Regarding the long (32) spaced items, this may be an indicator of where this condition gets its advantage, especially considering that recalled long (32) spaced word repetitions showed more theta power than all other conditions. The overall pattern of effects shows that when information is primed and in working memory it does not need to be retrieved (massed), and these items do not gain the repetition advantage of information that needs to be retrieved and processed (spaced).

In the voltage similarity analysis, a more similar neural state during the repetition (compared to the initial presentation) leads to better subsequent memory. This again support study-phase retrieval (spaced) and deficient processing (massed), as well as challenges contextual variability which would predict a decrease in similarity for spaced items. The time–frequency similarity results

were difficult to interpret, but they do not seem to help with the interpretation of the summarized results.

Overall, the spacing effect seems to be most consistently driven by the retrieval of a prior occurrence and additional semantic processing, and therefore it seems that study-phase retrieval is a key part of the spacing effect. Massed items do seem to be put at a disadvantage via deficient processing, though it seems likely that effects like these may only occur in a relatively unnatural experimental setting.

## **Chapter 5**

### **General discussion**

(nb: this is incomplete)

#### **5.1 Summary of results**

The results of the two experiments followed the same overall pattern. Based on the behavioral results of Experiment 2, we hoped that the additional spaced study lags would have provided more graded neural effects, but the analyses undertaken here do provide additional insights. To summarize at a high level, it seems like information that is no longer in working memory shows evidence of retrieval during a study repetition, and this benefits performance on a subsequent test.

The results across modalities (ERP, time–frequency, and similarity for each) are not directly comparable, but we should not evaluate them in a vacuum, either. We take a relatively conservative approach when integrating these results given the consistency of support for particular hypotheses, not to mention that assessing every pattern across modalities has the potential to explode combinatorically.

##### **5.1.1 ERP results**

The ERP effects in Experiment 1 and Experiment 2 aligned well overall, and there were no significant discrepancies. We expected to see some lag effects in Experiment 2, but unfortunately none of the ERP components followed the full gradient. The closest to a lag effect here was for LPC latency, where the peak occurred later for medium (12) and long (32) spaced compared to

massed repetitions; massed did not differ significantly from short (2) spaced repetitions, showing a separation between shorter and longer lags. The LPC is associated with conscious recognition of a stimulus, and it seems reasonable that primed items would be recognized faster than those whose representations are no longer in working memory.

The visual N1 ERP component, which is related to early attentional processes, showed that massed repetitions attenuated compared to spaced repetitions. While the effects were only marginal in Experiment 2, the overall pattern from the Experiment 1 persisted and should not be dismissed. (We note again that almost all ERP component voltages were attenuated in the second experiment, perhaps because of slightly different electrode net placements across the two sessions.)

The N400 ERP component, which is related to semantic processing, attenuates when the information being processed is in working memory, reflecting that the required semantic information is already activated.

The LPC, which is related to consciously accessing information, showed effects that align with an explanation that centers around working memory. This component was not described from a working memory perspective in previous analyses in repetition experiments (Olichney et al., 2000; Van Strien et al., 2007, rather, they discussed it as “template matching”), but the pattern of effects follows this idea.

### **5.1.2 Time–frequency results**

Effects in the theta band were consistent across experiments. Both showed sustained theta synchrony across spaced word repetitions and into image presentations compared to massed repetitions, and this was particularly true for long (32) spaced repetitions in Experiment 2. Because theta is involved in memory retrieval and encoding and that these effects are seen when word–image associations are being formed (participants are asked to associate the word and image), this leads to the idea that study-phase retrieval and re-encoding are occurring more for spaced repetitions compared to massed repetitions. This explanation makes sense because our previous ERP analyses have shown that massed items are still in working memory. Sustained theta synchrony at longer

lags denotes that information needs to be retrieved from long-term memory for these repetitions.

Regarding the negative theta SME for massed repetitions, because these are immediate repetitions it seems unlikely that there has been much contextual drift across the two presentations. If this effect is the result of a contextual difference, participants may have considered the stimuli differently on each presentation. We did not measure how participants perceived stimuli, and thus it is difficult to make strong claims about the negative SME.

#### Lower alpha

In lower alpha, the only difference between experiments was for image repetitions: Experiment 1 showed that massed images were more negative only in the first time window. Experiment 2 showed that spaced images had a larger power decrease from the first to the second time window compared to massed images, though massed were still lower overall. Perhaps the relative decrease during stimulus processing is an important factor in explanations of attentional effects, something that we expect to find a neural correlate for given that spaced items are remembered better than massed on average.

#### Upper alpha

In comparison to Experiment 1, upper alpha for massed items in Experiment 2 stayed desynchronized for longer

Lower beta effects were mainly the same across the two experiments. There was more desynchronization for subsequently remembered words and images (expected), implying that more semantic processing occurs.

### 5.1.3 Similarity results

Voltage similarity results maintained similar results across the experiments and support study-phase retrieval. This is because spaced items (regardless of lag in Experiment 2?) stay consistently similar across image presentations, leading to the idea that these items are retrieved from memory and processed in a similar way.

Time-frequency similarity results showed a complicated pattern of results, for which it is

difficult to distill an overall picture.

## **5.2 Alignment with the hypotheses**

### **5.2.1 Deficient Processing**

Under deficient processing, we would expect that

The attenuation of the visual N1 ERP component for massed compared to spaced repetitions implies that massed items may be processed at a level that is disadvantageous. This supports the deficient processing hypothesis.

### **5.2.2 Contextual Variability**

Regarding the negative theta SME for massed images, negative SMEs in the theta band have been induced when context is purposely varied (Staudigl & Hanslmayr, 2013). An idea approached earlier is that remembered massed image repetitions in both experiments may have had a more varied context (and a negative SME) because they were thought about in a different way. The time-frequency similarity results do not support this because the massed image study presentations increase in similarity across time. However, the similarity results for ERPs show that massed items decrease in similarity across time, though this is not specific to recalled items and this is also integrating across modalities which may not be a valid comparison.

### **5.2.3 Study-Phase Retrieval**

Previous research has shown theta to be important for memory processing. Fuentemilla, Penny, Cashdollar, Bunzeck, and Düzel (2010) showed that memory reactivation involves increased theta power, coupled with activity in the beta and gamma bands. The tasks in the present experiments have not been working memory tasks requiring the maintenance and manipulation of information over short periods of time, but the reactivation of memory representations during stimulus repetitions correlates with lag (and with behavioral performance); longer lags showed more theta power. Theta synchrony along the visual ventral stream (occipitotemporal region) is

important for maintaining cortical representations (Düzel, Penny, & Burgess, 2010), which is one reason why elevated levels of theta in the medial temporal lobes prior to stimulus presentation can lead to better memory encoding.

### 5.3 Future directions

For future analyses, comparing the similarity of each during test to its study presentations would be another way to examine contextual reinstatement (Howard & Kahana, 2002; Lohnas & Kahana, 2014). This analysis would have a bearing on study-phase retrieval and contextual variability. Contextual variability would be supported if a spaced test image is similar to only one study image, whereas study-phase retrieval would be supported if the test image was similar to some combination of the study images. This analysis would benefit from separating item and context features in neural patterns.

For future experiments, it would still be enlightening to get away from a paradigm that involves immediate repetition so that deficient processing could be eliminated as a possible theory. An interesting way to take this would be to use an established spacing effect paradigm from the literature (e.g., Cepeda et al., 2009, Experiment 1) and add EEG recordings. This would involve a multi-session experiment and have the added benefit of being more applicable to real-world learning. It may also simplify the possibilities for why the spacing effect occurs (deficient processing should essentially be eliminated).

## References

- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Skyes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1: Memory in everyday life, pp. 396–401). Oxford, England: Wiley & Sons.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–123). John Wiley & Sons.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, *10*(4), 433–436.
- Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: evidence from once-presented words. *Memory*, *6*(1), 37–65. doi: 10.1080/741941599
- Burke, J. F., Zaghoul, K. A., Jacobs, J., Williams, R. B., Sperling, M. R., Sharan, A. D., & Kahana, M. J. (2013). Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J Neurosci*, *33*(1), 292–304. doi: 10.1523/JNEUROSCI.2057-12.2013
- Callan, D. E., & Schweighofer, N. (2010). Neural correlates of the spacing effect in explicit verbal semantic encoding support the deficient-processing theory. *Hum Brain Mapp*, *31*(4), 645–659. doi: 10.1002/hbm.20894
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378.
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Exp Psychol*, *56*(4), 236–246. doi: 10.1027/1618-3169.56.4.236
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol Bull*, *132*(3), 354–380. doi: 10.1037/0033-295X.132.3.354
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 389–396.
- Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn Affect Behav Neurosci*, *2*(1), 1–18.
- Delaney, P. F., Verkoeijen, P. P., & Spiegel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 53, pp. 63–147). Burlington: Elsevier.
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627–634.

- Duarte, A., Ranganath, C., Trujillo, C., & Knight, R. T. (2006). Intact recollection memory in high-performing older adults: ERP and behavioral evidence. *J Cogn Neurosci*, *18*(1), 33–47. doi: 10.1162/089892906775249988
- Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Brain Research. Cognitive Brain Research*, *18*(3), 255–272.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4—58. doi: 10.1177/1529100612453266
- Düzel, E., Penny, W. D., & Burgess, N. (2010). Brain oscillations and memory. *Current Opinion in Neurobiology*, *20*(2), 143–149. doi: 10.1016/j.conb.2010.01.004
- Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychol Rev*, *62*(3), 145–154.
- Fellner, M.-C., Bäuml, K.-H. T., & Hanslmayr, S. (2013). Brain oscillatory subsequent memory effects differ in power and long-range synchronization between semantic and survival processing. *Neuroimage*, *79*, 361–370. doi: 10.1016/j.neuroimage.2013.04.121
- Friedman, D., & Johnson, R., Jr. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microsc Res Tech*, *51*(1), 6–28.
- Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-coupled periodic replay in working memory. *Curr Biol*, *20*(7), 606–612. doi: 10.1016/j.cub.2010.01.057
- Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 625–630.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16.
- Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(3), 282–294.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Mem Cognit*, *7*(2), 95–112.
- Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371–377.
- Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112.
- Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., & Scholte, H. S. (2012). Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Comput Biol*, *8*(10), e1002726. doi: 10.1371/journal.pcbi.1002726
- Hanslmayr, S., Spitzer, B., & Bäuml, K.-H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cereb Cortex*, *19*(7), 1631–1640. doi: 10.1093/cercor/bhn197
- Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories—a processing based perspective on oscillatory subsequent memory effects. *Neuroimage*, *85 Pt 2*, 648–655. doi: 10.1016/j.neuroimage.2013.05.121
- Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front Hum Neurosci*.

- Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bäuml, K.-H. T. (2011). The relationship between brain oscillations and BOLD signal during memory formation: a combined EEG-fMRI study. *J Neurosci*, *31*(44), 15674–15680. doi: 10.1523/JNEUROSCI.3140-11.2011
- Henson, R. N., Rylands, A., Ross, E., Vuilleumeir, P., & Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage*, *21*(4), 1674–1689. doi: 10.1016/j.neuroimage.2003.12.020
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium* (pp. 77–97). Lawrence Erlbaum Associates.
- Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*(1), 70–74.
- Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(1), 31.
- Howard, M., & Kahana, M. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.
- Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 153–160.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Mem Cognit*, *24*(1), 103–109.
- Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychon Bull Rev*, *12*(1), 159–164.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*, 141–151. doi: doi:10.1177/001316446002000116
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J Exp Psychol Learn Mem Cogn*, *33*(4), 704–719. doi: 10.1037/0278-7393.33.4.704
- Khajah, M. M., Lindsey, R. V., & Mozer, M. C. (2014). Maximizing students' retention via spaced review: Practical guidance from computational models of memory. *Top Cogn Sci*, *6*(1), 157–169. doi: 10.1111/tops.12077
- Kim, M., Kim, J., & Kwon, J. S. (2001). The effect of immediate and delayed word repetition on event-related potential in a continuous recognition task. *Brain Res Cogn Brain Res*, *11*(3), 387–396.
- Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev*, *29*(2-3), 169–195.
- Klimesch, W., Doppelmayr, M., Russegger, H., & Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. *Neuroreport*, *7*(7), 1235–1240.
- Klimesch, W., Hanslmayr, S., Sauseng, P., Gruber, W., Brozinsky, C. J., Kroll, N. E. A., ... Doppelmayr, M. (2006). Oscillatory eeg correlates of episodic trace decay. *Cereb Cortex*, *16*(2), 280–290. doi: 10.1093/cercor/bhi107
- Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res Rev*, *53*(1), 63–88. doi: 10.1016/j.brainresrev.2006.06.003
- Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations. *Exp Psychol*, *52*(2), 99–108.
- Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., & Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the p1-n1 complex and are related to memory per-

- formance. *Brain Res Cogn Brain Res*, *19*(3), 302–316. doi: 10.1016/j.cogbrainres.2003.11.016
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci*, *2*, 4. doi: 10.3389/neuro.06.004.2008
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*, *62*, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.
- Lega, B. C., Jacobs, J., & Kahana, M. (2011). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*. doi: 10.1002/hipo.20937
- Lohnas, L. J., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *J Mem Lang*, *64*(3), 249–255. doi: 10.1016/j.jml.2010.11.003
- Long, N. M., Burke, J. F., & Kahana, M. J. (2013). Subsequent memory effect in intracranial and scalp eeg. *Neuroimage*, *84C*, 488–494. doi: 10.1016/j.neuroimage.2013.08.052
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn Sci*, *4*(11), 432–440.
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *J Exp Psychol Learn Mem Cogn*, *31*(2), 322–336. doi: 10.1037/0278-7393.31.2.322
- Mammarella, N., Russo, R., & Avons, S. E. (2002). Spacing effects in cued-memory tasks for unfamiliar faces and nonwords. *Mem Cognit*, *30*(8), 1238–1251.
- Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc Natl Acad Sci U S A*, *108*(31), 12893–12897. doi: 10.1073/pnas.1015174108
- Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*(3), 530–540. doi: 10.1016/j.neuron.2007.08.017
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, *158*(3800), 532. doi: 10.1126/science.158.3800.532-b
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606.
- Mensink, G.-J. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*(2), 172–186.
- Mitchell, D. J., McNaughton, N., Flanagan, D., & Kirk, I. J. (2008). Frontal-midline theta from the perspective of hippocampal “theta”. *Prog Neurobiol*, *86*(3), 156–185. doi: 10.1016/j.pneurobio.2008.09.005
- Mozer, M. C., Pashler, H., Cepeda, N., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In *Advances in neural information processing systems* (pp. 1321–1329).
- Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1023–1035. doi: 10.1016/j.neubiorev.2009.12.014

- Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia: Electrophysiological measures of impaired and spared memory. *Brain*, *123* ( Pt 9), 1948-1963.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 156869. doi: 10.1155/2011/156869
- Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: choices and consequences. *Psychonomic Bulletin and Review*, *14*(2), 187–193.
- Pavlik, P. I., Jr, & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cogn Sci*, *29*(4), 559–586. doi: 10.1207/s15516709cog0000\_14
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*(11), 1842–1857.
- Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (Vol. 22, pp. 1090–1104). IEEE.
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. doi: 10.1016/j.tics.2005.12.004
- Poldrack, R. A., & Wagner, A. D. (2004). What can neuroimaging tell us about the mind? Insights from prefrontal cortex. *Current Directions in Psychological Science*, *13*(5), 177–181.
- Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, *12*(1), 24–30. doi: 10.1016/j.tics.2007.10.010
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychol Rev*, *116*(1), 129–156. doi: 10.1037/a0014420
- Proverbio, A. M., & Adorni, R. (2009). C1 and P1 visual responses to words are enhanced by attention to orthographic vs. lexical properties. *Neurosci Lett*, *463*(3), 228–233. doi: 10.1016/j.neulet.2009.08.001
- Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452.
- Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *J Exp Psychol Learn Mem Cogn*, *24*(1), 161–172.
- Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychol Rev*, *115*(4), 893–912. doi: 10.1037/a0013396
- Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *J Neurosci*, *23*(34), 10809–10814.
- Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Curr Biol*, *23*(12), 1101–1106. doi: 10.1016/j.cub.2013.04.074
- Su, L., Fonteneau, E., Marslen-Wilson, W., & Kriegeskorte, N. (2012). Spatiotemporal search-light representational similarity analysis in EMEG source space. In *Pattern recognition in neuroimaging (prni), 2012 international workshop on* (pp. 97–100).
- Summerfield, C., & Mangels, J. A. (2005). Coherent theta-band EEG activity predicts item-context binding during encoding. *NeuroImage*, *24*(3), 692–703. doi: 10.1016/j.neuroimage.2004.09.012

- Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536.
- Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*(3), 154–163.
- Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352–373.
- Turk-Browne, N. B., Simon, M. G., & Sederberg, P. B. (2012). Scene representations in parahippocampal cortex depend on temporal context. *J Neurosci*, *32*(21), 7202–7207. doi: 10.1523/JNEUROSCI.0942-12.2012
- Van Strien, J. W., Verkoeijen, P. P. J. L., Van der Meer, N., & Franken, I. H. A. (2007). Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. *Int J Psychophysiol*, *66*(3), 205–214. doi: 10.1016/j.ijpsycho.2007.07.003
- Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, *1122*(1), 161–170. doi: 10.1016/j.brainres.2006.09.023
- Vogel, E. K., & Luck, S. J. (2000). The visual n1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.
- Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: when priming hinders new episodic learning. *J Cogn Neurosci*, *12 Suppl 2*, 52–60. doi: 10.1162/089892900564064
- Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology*, *35*(1), 81–87.
- Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, *119*, 889–905.
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on* (pp. 3485–3492).
- Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*(6000), 97–101. doi: 10.1126/science.1193125
- Xue, G., Mei, L., Chen, C., Lu, Z.-L., Poldrack, R., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *J Cogn Neurosci*, *23*(7), 1624–1633. doi: 10.1162/jocn.2010.21532