Attentional fluctuations and the temporal organization of memory:

Insights from behavioral and pupillometry measures

Manasi Jayakumar

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

Attentional fluctuations and the temporal organization of memory:

Insights from behavioral and pupillometry measures

Manasi Jayakumar

Fluctuations in attention are ubiquitous. We all experience the waxing and waning of our attention, with periods of focus alternating with periods of distraction by irrelevant thoughts or external sensations. Given the pervasiveness of these fluctuations, it is important to understand how they influence both our behavior in the moment and the structure of our memory. In this dissertation, I use behavioral studies and eye tracking to measure spontaneous attentional fluctuations and examine how these fluctuations shape online behavior and subsequent memory. I test my primary hypothesis that optimal attentional states help us link experiences over time to allow our memories to be temporally organized, whereas suboptimal states disrupt the temporal structure of memory. In Chapter 1, I present four studies using a novel experimental design to connect research on sustained attention and memory. I replicate prior findings linking response-time-based measures of attention to online behavior. Surprisingly, I found that these response-time measures of attention do not predict the temporal structure of free recall. In Chapter 2, I indexed attentional fluctuations with both response times and pre-trial pupil size and demonstrated that these measures of attention predict complementary aspects of behavior. Attentional fluctuations, as indexed by pupil size, predicted the temporal organization of memory but not attentional lapses in online behavior. Conversely, response times predicted attentional lapses in the moment but did not predict the temporal organization of memory. Finally, in ongoing work in Chapter 3, I examine whether providing cues at retrieval enhances or diminishes the effects of attentional fluctuations on the temporal organization of memory. Together, my results shed light on the complex interactions between fluctuations in attention and episodic memory. Critically, I show that different measures of attention – behavioral vs. physiological approaches – capture distinct aspects of cognitive function, and suggest that the attentional states that shape online behavior and later memory are at least partly distinct.

Table of Contents

iv

List of Figures

List of Tables

Acknowledgments

To my advisor, Mariam Aly: I truly admire your wealth of knowledge and your attention to detail at every stage of research. I am proud of the work we've done together and I am so grateful for your unwavering support and guidance through the last 6 years. Thank you for shaping me into the researcher and scientist I am today.

To my secondary advisor, Lila Davachi: I truly admire your enthusiasm and passion for science and love of looking at the big picture. Thank you for always cheering me and reminding me to believe in myself.

To co-PI of the Alyssano group, Chris Baldassano: I truly admire your creativity with ideas, methods, and analyses. Thank you for your mentorship and support and always being available when I needed advice. I truly appreciated your generosity and kindness during my time at Columbia.

To my committee members, Mike Esterman and Daphna Shohamy: Thank you for your knowledge, time, and feedback. I am so honored to have you both on my committee and to discuss my work with you.

To my research assistants, Aria, Cecilia, Chinmayi, Emily, and Raheema: Thank you so much for all your help and support - none of the work in this dissertation would have been possible without your efforts.

To Caroline Marvin and Katherine Fox-Glassman: Thank you for your continued support and encouragement. I would not be who I am today without you both!

To Erin, Jon, Joanna, Kola, and Maria: Thank you! You make this all possible from behind the scenes.

To CTL staff, Ian, Abby, Caitlin, Mark, Glen, Nicky, and others I've met over the years: Thank you for creating a supportive space that allowed me to discover myself amongst like-minded peers and colleagues. Thank you for opening the doors to a dream I did not even realize I had.

To my former mentors and mentors outside of Columbia: Thank you for taking a chance on me when I was just starting off. I would not be here today without your patience, support, and mentorship.

To past and present Aly lab and Alyssano group members, Alex, Aria, Caroline, Chey, Craig, Eren, Halle, Hannah, Jiawen, Marta, Matt, Mike, Narjes, Nick, Raheema, Sam, Serra, Taylor, and Zall: Thank you for your collaboration, support, and friendship! You made grad school fun!

To former grad students, Paul Bloom and Monica Thieu: Thank you for your advice and support on all things R, statistics, pedagogy, and academia!

To fellow department grad students, postdocs, and lab managers: I thank you for your support, friendship, and laughter over the years! There are so many of you that I'll forget to name a few of you – but you all know who you are.

To the mental health professionals I've seen through the years: Thank you for helping me develop the skills and the strength to get through the last 6 years of ups and downs!

To Hannah and Halle: Thank you for the support and love through the ups and downs of the last 6 years (and counting)!

To Caroline: Thank you for always being there for me!

To my friends around the world: thank you for keeping me sane and helping me see the lighter side of things. Again, there are so many of you and I'm afraid I'd forget someone. If you've had contact with me during my time in grad school, you are included on this list!

To my extended family: Thank you for supporting me on this journey that has lasted 10 years.

To all the wonderful puppers I've met over the years, including Izzie, Theo, Boba, Sophie, the Halloween spider-dogs, and Willie: Thank you for the emotional regulation and stress relief.

To Navya: Thank you for checking in when I disappear for months at a time. Thank you for always being there for me!

To Hanumanth and Ramya: Thank you for your support and encouragement.

To Santhi Amma: Thank for you being a second mother to me and for ensuring I take care of myself.

To Zuko and Tuco: Thank you for all the popcorns, cuddles, and panic vet visits!

To Amma and Appa: I am who I am because of you both. Thank you for always supporting me and loving me.

To Soma: Words cannot express my gratitude. Thank you for everything!

Dedication

To Amma, Appa, and Soma – the three pillars of my life.

To Karthik – whose presence is felt in everything I do.

Introduction

Memory is temporally organized: recalling a specific event often acts as a cue for recall of other events that occurred close to it in time (Howard & Kahana, 2002a). This phenomenon is extremely robust and is a consistent feature of episodic memory. However, it can be influenced by both internal and environmental factors (Healey et al., 2019). In this dissertation, I examine attentional fluctuations as one of the factors that could influence the temporal organization of episodic memory.

Fluctuations in attention are ubiquitous. As we experience our daily life, attention waxes and wanes with periods of focus interrupted by periods of distractions and irrelevant thoughts (Smallwood & Schooler, 2006; Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2015). Given the pervasiveness of these spontaneous shifts in attention, it is important to examine how they may impact subsequent memory organization and retrieval. Yet, prior to my dissertation work, the influence of these attentional fluctuations on the temporal organization of episodic memory had not been explored. Thus, this dissertation examines the influence of attentional fluctuations on the temporal structure of memory. Chapters 1-3 describe three studies that I conducted to address this goal. This Introduction section reviews prior research and lays the conceptual and empirical foundation for Chapters 1-3.

In this section, I first review research on the temporal organization of memory, including proposed computational and theoretical frameworks and methods of quantifying the temporal structure of memory. Next, I briefly describe the rich history of work on attention and memory, as well as how it ties into this dissertation. I follow this up with brief reviews of experimental designs to capture fluctuations in attention, and empirical work using behavioral and pupil-based

measures of attention. Finally, I provide an overview of the novel research conducted in this dissertation.

Temporal Organization of Memory

(Howard & Kahana, 2002a) introduced a computational model - the Temporal Context Model (TCM) - to account for the robust temporal organization observed in episodic memory. As proposed by the model, at initial encoding, item representations are linked to a slowly evolving temporal context representation. At retrieval, recall of one of the items triggers the reinstatement of the temporal context and subsequently the recall of other items encoded nearby in time. This results in items being recalled together if they were encoded close together in time. (Howard & Kahana, 2002a; Polyn & Kahana, 2008). While TCM was introduced to specifically explain the temporal organization in retrieval, a later version of the model (the Context Maintenance and Retrieval Model or CMR), also incorporates semantic and source information as potential factors shaping recall (Polyn et al., 2009).

An often-used approach to studying the temporal organization of memory is the free recall paradigm, where participants encode a list of items and are later asked to recall as many items as they can in any order they can (Kahana, 1996; Polyn et al., 2009). Lag-Conditional Response Probability (lag-CRP) curves provide the probability of recalling two items successively, given their distance at encoding (e.g. Kahana, 1996; Howard & Kahana, 2002; Palombo et al., 2019; Diamond & Levine, 2020; see Healey et al., 2019 for a review). The distance at encoding or lag is typically plotted on the x-axis as the number of intervening encoding-list items between the two successively recalled items. The conditional-response probability (CRP) is plotted on the y-axis. Given the recall of an item at position i during encoding, the Conditional Response Probability is the likelihood of recalling another item at a

specific lag from i, conditional on the number of opportunities to make a recall at that particular lag.

These lag-CRP curves show two characteristic properties: 1) temporal contiguity and 2) forward asymmetry. The temporal contiguity effect (also known as the temporal clustering effect), is the tendency to cluster items at recall such that items encoded close in time are recalled close in time. The forward asymmetry effect is the tendency to recall items in a forward direction (the direction of initial encoding) compared to a backward direction (opposite to the direction of initial encoding; Kahana, 1996; Healey et al., 2019). This may seem intuitive when you relate it to your daily life: at the end of a regular working day, you are likely to remember things in order from the start to the end of the day (i.e. in the forward direction) such as waking up, drinking coffee, going to work, and so on.

Many factors can influence the temporal organization of recall - some of these can be cognitive factors, while others may be environmental. Research has shown that temporal organization is influenced by both experimentally-manipulated factors such as the modality of stimuli presentation (auditory vs. visual), or the length of lists (Healey et al., 2019), as well as other participant-specific factors, such as the presence of certain psychiatric conditions (e.g. schizophrenia or psychosis; Polyn et al., 2015; Murty et al., 2018). One interesting factor that has been shown to influence the temporal organization of memory is event segmentation. Event segmentation theory proposes that continuous experiences in our day-to-day life are segmented into separate events, with the changes in context between these event segments serving as event boundaries (Zacks et al., 2007; Clewett & Davachi, 2017). Importantly, memory retrieval is shaped by these event boundaries; when items are encoded within the same event, there is better order memory, and the items are remembered as being closer together (DuBrow & Davachi,

2013, 2016; Heusser et al., 2018; Clewett et al., 2020). Furthermore, these event boundaries can be shifts in our internal cognitive states(Y. C. Wang et al., 2023).

Thus, temporal context models and the resulting research provide a foundational framework for understanding the temporal organization of memory. Understanding the factors that interact with temporal context to influence memory organization can provide valuable insights into human episodic memory.

Attention and Memory

Attention and episodic memory have a robust relationship that has been extensively studied; attention towards a specific feature of an item or event improves later memory for that item or event (see Chun & Turk-Browne, 2007 and Aly & Turk-Browne, 2017 for reviews). Particularly relevant to this dissertation is the work on divided attention at encoding. These studies demonstrate a stark reduction in retrieval performance when attention is divided at encoding using a dual task; this is observed across free recall, cued recall, recognition, and associative memory tests (Craik et al., 1996; Anderson et al., 1998; Troyer et al., 1999; Craik et al., 2018). Importantly, divided attention at encoding also influences temporal order memory, with a dual-task at encoding resulting in worse memory for temporal order (Troyer & Craik, 2000). However, one limitation of these divided attention tasks is that attention is manipulated experimentally, and thus, the resulting effects may not be comparable to those from more subtle and spontaneous attentional fluctuations that we experience in daily life. Nevertheless, the research on divided attention provides a solid foundation that differences in attention at encoding indeed influence subsequent memory as well as its organization. Building on this, I now turn to experimental designs that aim to capture endogenous fluctuations in attention.

Capturing Spontaneous Fluctuations in Attention

Two bodies of work are relevant to how spontaneous fluctuations in attention can be examined. First is the work on mind-wandering that aims to capture the subjective experience of drifting away from a task towards unrelated thoughts (Smallwood & Schooler, 2006). These studies typically use the probe method of experience sampling where participants are intermittently interrupted during a task and asked to report their subjective experience at that time (Smallwood & Schooler, 2015). Participants may be probed to respond if they were "on task" or "mind-wandering" (e.g., Metcalfe & Xu, 2016; Xu & Metcalfe, 2016; Xu et al., 2018; Garlitch & Wahlheim, 2020); some tasks even have participants break down the mind-wandering report into different types such as external vs. internal or task-related vs. task-unrelated (e.g., Unsworth & Robison, 2016, 2018). Other studies ask participants to describe what they were thinking right before the probe was presented and later code for the details in the report (e.g. Smallwood et al., 2003). Thus, these thought probes provide an assay of attentional fluctuations with participants being "on task" for some trials and "mind-wandering" for others.

Examining the relationship between these mind-wandering states and subsequent memory, studies typically find that mind-wandering or "off-task" periods are associated with worse memory performance (Smallwood et al., 2003; Risko et al., 2012; Garlitch & Wahlheim, 2020; Martarelli & Ovalle-Fresa, 2021). However, the one limitation of these studies is that thought probes only capture attention at a few discrete time points during the task. Given that this method doesn't provide us with a way of capturing the attentional state for each trial, it would not be possible to associate these states with temporal organization at recall.

The second body of work relevant to capturing spontaneous fluctuations in attention is the research on sustained attention; this body of research provides a way of measuring moment-

by-moment fluctuations in attention during a task (Esterman & Rothlein, 2019). Some of the earliest work in this line of research comes from studies of vigilance, undertaken to understand how fatigue influenced radar operators during World War II (Rosenberg et al., 2013). These vigilance tasks involved detecting a rare and unpredictable event, mirroring the nature of the task that the radar operators performed (waiting for a rare event to be detected on the radar; Robertson et al., 1997). Thus, these vigilance tasks had a similar limitation as the mind-wandering tasks in that responses were restricted to intermittent and discrete time points, thus failing to provide estimates of attentional quality on a moment-by-moment basis (Rosenberg et al., 2013).

In contrast, a subset of tasks called not-X Continuous Performance Tasks (not-X CPTs) were developed: here, participants are asked to respond to a frequent event but withhold the response for the rare event (Rosenberg et al., 2013), essentially performing a go/no-go task. A novel version of the not-X CPT is the gradual onset continuous performance task (gradCPT) introduced by Esterman et al., (2013). This task introduces a gradual transition between the images presented on adjacent trials, such that one image fades into the other. This helps to prevent bottom-up attentional capture common with abrupt onset of stimuli. Importantly, both the not X-CPTs and gradCPT provide moment-by-moment measures of attentional state. Therefore, gradCPT is the optimal task for capturing spontaneous attentional fluctuations and I introduce a modified version in Chapter 1.

Measuring Attention with Response Times

Sustained attention tasks, like the gradCPT, use response times to measure attentional states on a moment-by-moment basis. In the gradCPT, Response Time (RT) variability is used to index attentional states. Good attentional states ("in the zone") are characterized by the lower RT

variability and worse attentional states ("out of the zone") are characterized by the higher RT variability. A robust finding from these studies is the difference in online task performance by attentional state: participants consistently make more errors during "out of the zone" periods (Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014; Rosenberg et al., 2017; Fortenbaugh et al., 2018; Esterman & Rothlein, 2019).

A second commonly used approach is to compute the average RT of the three preceding trials to measure the attentional state for an item (e.g., Robertson et al., 1997; deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022). Here, faster RTs in the preceding trials (indicative of habitual responding) are characterized as the worse attentional state, while slower RTs form the better attentional state. As in the work with RT variability, the three preceding trials approach has been shown to predict online task errors (deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022).

In addition to RT-based measures predicting online task performance, they have also been linked to subsequent memory (e.g., deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022). In these studies, old/new recognition tests were used to assess memory for the encoded scenes; good attentional states were found to be associated with better memory performance. However, these studies do not test for the temporal organization of memory.

Chapter 1 of this dissertation builds on these studies by indexing attentional states moment-by-moment at encoding, and examining their influence on temporal organization at recall. Chapters 2 and 3, while focusing on pupil size as a primary measure, also use RT-based measures of attention as a secondary measure.

Pupil Measures of Attention

There is a rich history of using pupil measures as an indicator of attention and arousal (Aston-Jones & Cohen, 2005; Joshi & Gold, 2020; Laeng et al., 2012; Mathot, 2018; Strauch et al., 2022). While the earliest work demonstrated that pupil size could be used as an assay of emotional arousal (Hess & Polt, 1960, 1964), subsequent work demonstrated its potential as an indicator of alertness (Lowenstein et al., 1963), mental effort or processing demands (Kahneman & Beatty, 1966; Kahneman, 1973; Beatty, 1982; Richer & Beatty, 1987) and decision-making (de Gee et al., 2014). Two measures of pupil size have been consistently used as a measure of arousal and attention: the baseline pupil size and evoked pupil responses.

The baseline pupil size involves the measurement of the pupil size during the 500-1000ms leading up to the onset of a trial's stimulus; this provides a measure of the internal state just prior to stimulus onset. The evoked pupil response examines the pupil response to the onset of the stimulus in terms of the amplitude of change from baseline until its return back to baseline. (Aston-Jones & Cohen, 2005; Sirois & Brisson, 2014; Joshi & Gold, 2020).

These measures of pupil size predict behavioral and cognitive measures of attentional lapses. For instance, smaller baseline pupil sizes have been associated with lapses in attention (van den Brink et al., 2016; Unsworth et al., 2020), as have smaller phasic responses (Unsworth et al., 2018). Other studies have found that "off-task" reports on the mind-wandering thought probes were associated with smaller baseline pupil size and smaller phasic responses (Unsworth & Robison, 2016, 2018). Pupil size has also been related to subsequent memory (e.g., Kafkas & Montaldi, 2011; Otero et al., 2011; Papesh et al., 2012; Rijn et al., 2012; Naber et al., 2013; Ariel & Castel, 2014; Whitlock et al., 2023). Larger evoked pupil size during encoding has been associated with better recognition memory (e.g., Kafkas & Montaldi, 2011; Papesh et al., 2012;

Naber et al., 2013; Ariel & Castel, 2014; Hoffing & Seitz, 2014) and better associative memory (Whitlock et al., 2023). However, none of these studies examine the relationship between pupil size and the temporal organization of memory at retrieval.

Relevant to the current dissertation, one study has examined the relationship between evoked pupil responses and event boundaries (Clewett et al., 2020). The study found that pupil dilations track event boundaries and that these boundary-related signals are related to subsequent memory for temporal recency. Using a dimensionality reduction technique, the study found that an early-peaking component of the evoked pupil response was related to temporal order memory. Furthermore, an exploratory analysis found that the stability of the pupil size between two items at encoding was related to subsequent performance on temporal recency judgments, suggesting that pupil measures at encoding may indeed be related to the incorporation of temporal information in memory (Clewett et al., 2020).

Chapters 2 (and 3) of this dissertation build on these prior findings and studies. First, while prior work on pupil size and memory has focused extensively on evoked pupil responses, we used baseline pupil size given the strong relationship between this measure and attentional lapses, as described above. This allows us to index attentional states leading up to the trial onset, thus focusing on a pupil measure of the internal cognitive state independently from the response to the trial or the appearance of a visual stimulus. Second, we used free recall and the associated lag-CRP approach to quantify the temporal organization of recall. Lag-CRPs provide us with further insight into the temporal dynamics of recall by quantifying temporal contiguity and forward asymmetry effects; these effects can then be explored as a function of baseline pupil size at encoding.

Overview of Dissertation

To summarize, attention and memory are tightly intertwined, with attention being an important factor in predicting subsequent memory retrieval and its temporal organization. While prior research has shown a robust relationship between experimenter-manipulated attention and memory, the question of how endogenous fluctuations in attention influence memory is underexplored. Thus, in this dissertation, I examined how spontaneous fluctuations in attention shape online behavior and subsequent memory, using behavioral studies and pupillometry. Across the three chapters in this dissertation, I hypothesized that optimal attentional states sustain temporal organization of recall while suboptimal attentional states disrupt them.

In Chapter 1, I introduced a novel experimental design that connected research on sustained attention and episodic memory. Using response-time-based measures of attentional states, I replicated prior research showing more behavioral errors during periods of worse attention. Interestingly, I did not find any effects of attentional states on the temporal organization of free recall: recall was strongly temporally organized regardless of attentional state at encoding. This suggests that either the temporal organization of memory is robust to attentional fluctuations, or that RT does not capture the kinds of attentional fluctuations that influence the temporal organization of memory.

I tested these alternative hypotheses in Chapter 2, by using two complementary measures of attentional states: response times and baseline pupil size. I found that attentional fluctuations, as indexed by pupil size, predicted the temporal organization of recall; however, pupil size did not predict online task performance. On the other hand, response-time-based measures of attentional state predicted online task performance, but not the temporal organization of recall. These results provide two critical conclusions. First, attentional fluctuations (as indexed by pupil

size) shape the temporal organization of memory. Furthermore, behavioral vs. physiological measures of attention predict different aspects of cognition and may be indexing distinct components of attention.

In ongoing work in chapter 3, I examine how retrieval cues interact with the attentional state at encoding to influence the temporal organization of memory. Once more, I index attentional fluctuations using both pupil size and response times. Preliminary results show no effect of pupil size on the temporal structure of memory when cues are provided at the test. However, given the low number of participants, it is too early to interpret the results from this work.

Together, my work in this dissertation highlights the complex interactions between attentional states and subsequent memory. It also contributes to our understanding of factors that influence the temporal organization of memory. More broadly, my work underscores the importance of using multiple, complementary measures to characterize attention and human behavior.

Chapter 1: Attentional fluctuations and the temporal organization of recall

Manasi Jayakumar, Chinmayi Balusu, Mariam Aly

1.1 Introduction

Episodic memories are temporally organized. Recall of a given event acts as a cue that can lead to recall of other events that were encoded close to it in time (Kahana, 1996; Howard & Kahana, 2002a; Healey et al., 2019). This is thought to occur because events encoded close to one another in time have similar internal "temporal context" representations (Howard & Kahana, 2002a). Although temporal contexts powerfully shape recall, the factors that drive how those temporal contexts form during encoding are relatively underexplored. A better understanding of those mechanisms can help individuals promote factors that create strong temporal contexts at encoding and thus enhance memory retrieval. Here, we test the hypothesis that natural fluctuations in attention during encoding contribute to temporal context representations.

The Temporal Context Model (TCM) was proposed to explain the temporal organization of episodic memory (Howard & Kahana, 2002a). According to this model, at encoding, item representations are linked to a slowly changing but ever-present temporal context. When an item is recalled, the temporal context from encoding is also retrieved. Using the retrieved temporal context as a cue, other items studied with an overlapping temporal context can then be recalled (Howard & Kahana, 2002a; Polyn & Kahana, 2008).

Free recall studies examining these effects typically use lag-Conditional Response Probability (lag-CRP) curves to characterize the temporal organization of recall (Kahana, 1996; Howard & Kahana, 2002a; Healey et al., 2019; Palombo et al., 2019; Diamond & Levine, 2020). These lag-CRP curves exemplify two characteristic features of temporally structured recall: the temporal contiguity effect and the forward asymmetry bias. The temporal contiguity effect refers to the tendency for items encoded close in time to be recalled close together (Howard & Kahana, 1999; Healey et al., 2019). The forward asymmetry bias refers to the higher probability of

successively recalling items in a forward vs. backward direction. That is, for a given recalled item, subsequently recalled items are more likely to have been encoded after (rather than before) the first-recalled item (Polyn & Kahana, 2008; Polyn et al., 2009. This forward asymmetry bias is thought to arise because a given item becomes part of the temporal context for succeeding items, and can thus serve as a memory cue.

This research has largely proposed that temporal contexts are ever-present and slowly drift during an encoding experience, as newly encountered items and thoughts are incorporated into the temporal context representation. Whether certain cognitive and environmental factors can affect the drift of temporal context is an important question, because such factors may promote (or hinder) memory retrieval (DuBrow et al., 2017). Recent research has shown that event segmentation can serve as one such factor that impacts the temporal organization of memory. Event segmentation theory hypothesizes that our ongoing experience is parsed into "segments" with the transition between segments acting as an event boundary (Zacks et al., 2007; Clewett & Davachi, 2017; Heusser et al., 2018). Memory retrieval is strongly shaped by such event boundaries: the order of events is remembered better, and events are more likely to be remembered as being closer together, if they were experienced within the same event segment vs. different event segments (DuBrow & Davachi, 2013, 2016; Heusser et al., 2018; also see Ezzyat & Davachi, 2010; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014).

We propose that similar to event boundaries, natural fluctuations in attention may contribute to how temporal contexts form during encoding. Fluctuations in attention are inherent aspects of human nature (Killingsworth & Gilbert, 2010). People tend to experience times when attention peaks, leading to intense focus on the task at hand, while at other times, attention wanes and focus is broken by intrusive thoughts, distractions, or fatigue (Smallwood et al., 2008).

14

Despite the ubiquity of such attentional fluctuations, past work examining how attention affects memory has typically focused on experimental manipulations of attention rather than spontaneous fluctuations. Such work has shown that experimental manipulations (such as directing participants' external attention towards a specific object, color, or spatial location) improve later memory for the attended event, but hurt memory for unattended events (Craik et al., 1996; Anderson et al., 1998; Naveh-Benjamin et al., 1998; Troyer et al., 1999; Troyer & Craik, 2000; Yi & Chun, 2005; Chun & Turk-Browne, 2007; Chun et al., 2010; Uncapher et al., 2011; Turk-Browne et al., 2013; LaRocque et al., 2015; Craik et al., 2018; see Aly & Turk-Browne, 2017).

A different body of work, on mind-wandering, has examined natural fluctuations of attention (Smallwood & Schooler, 2006; Christoff et al., 2016; also see Smallwood & Schooler, 2015). In these studies, participants are asked, with intermittent probes, to self-report whether they were "on task" or "mind-wandering" (e.g., Metcalfe & Xu, 2016; Xu & Metcalfe, 2016; Xu et al., 2018; Garlitch & Wahlheim, 2020) or asked to describe their thoughts at the time of the probe (Smallwood et al., 2003). These fluctuations in attention impact subsequent memory, such that mind wandering is related to worse memory (Smallwood et al., 2003; Risko et al., 2012; Garlitch & Wahlheim, 2020; Martarelli & Ovalle-Fresa, 2021). However, these studies are limited in that they only capture participants' attentional state at a few discrete time points and cannot precisely characterize the temporal dynamics of intrinsic fluctuations in attention. Furthermore, such studies have not determined how mind wandering may affect the temporal structure of memory.

Sustained attention research offers a way to measure moment-by-moment fluctuations in attention (Robertson et al., 1997; Sarter et al., 2001; Smallwood & Schooler, 2006; Esterman et

al., 2013; Rosenberg et al., 2013; deBettencourt et al., 2015; Rosenberg et al., 2017; deBettencourt et al., 2018; Fortenbaugh et al., 2018; Esterman & Rothlein, 2019; Decker et al., 2022; Elshiekh & Rajah, 2021; Wakeland-Hart et al., 2022). One type of task, the gradual onset continuous performance task (gradCPT), uses reaction time (RT) variability to index moment-by-moment fluctuations: trials with higher RT variability constitute "out of the zone" attentional states, while trials with lower RT variability reflect "in the zone" attentional states. Online task performance differs based on these attentional states: participants make more errors in the task during an "out of the zone" attentional state (Esterman et al., 2013, 2014; Rosenberg et al., 2013, 2017; Fortenbaugh et al., 2018). These studies, however, do not relate attentional fluctuations during the task to subsequent memory (see Madore et al., 2020 for a trait-level analysis). A related approach to characterizing attentional fluctuations based on RT showed that being in a good attentional state results in better recognition memory later on (deBettencourt et al., 2018; Wakeland-Hart et al., 2022), but these studies did not examine the temporal organization of memory. Building on these studies, in the current work, we use response time variability to characterize "in the zone" and "out of the zone" attentional states. We employed a modified version of the gradCPT, in which participants encoded trial-unique objects in the study phase (i.e., the sustained attention phase) and were later asked to verbally recall as many objects as they could in any order they chose. This allowed us to examine how moment-by-moment fluctuations in attention during encoding influence the temporal structure of recall.

We tested two main hypotheses. First, that recall will be more temporally structured when items are encoded "in the zone" vs. "out of the zone". This may occur because being focused on the task at hand leads to a consistent mental state, which can serve to bind and maintain a temporal context representation. Such focused attention may also facilitate linking

attended events in the environment to these internal temporal contexts. Conversely, the reduction of task-focused thought in a "bad" attentional state may result in switches between internally and externally focused thoughts that disrupt a consistent temporal context representation, and/or hurt the ability to link items in the environment to internal temporal contexts. If this is the case, then the hallmarks of temporally organized memory — temporal contiguity and forward asymmetry — may be enhanced for items encoded "in the zone" vs "out of the zone".

Our second hypothesis is motivated by the finding that recall can "leap" between cognitively similar but temporally distant events (Chan et al., 2017). This may occur because one's thoughts become integrated into temporal context representations: similar thoughts trigger similar temporal context representations, and in that way facilitate successive recall of items associated with cognitively similar contexts. We hypothesize that "in the zone" states, even if separated in time, constitute cognitively similar events: they consist of a focused mindset and the particular strategies that a person brings to mind to succeed in the ongoing task. On the other hand, we hypothesize that temporally distant "out of the zone" states are cognitively *dissimilar* events: every time an individual is unfocused, they may be unfocused in a different way, as their attention switches between the task and other ongoing, fluctuating thoughts. If this is the case, then recall may be more likely to "leap" between different "in the zone states" than different "out of the zone" states, bypassing items that were encoded in the other attentional state.

In sum, we aim to examine the behavioral effects of spontaneous attentional fluctuations on the temporal organization of recall. We characterize attentional fluctuations based on response time variability during an encoding task, classifying trials into relatively good "in the zone" and relatively worse "out of the zone" attentional states. To do that, we use a modified version of the gradCPT, introducing changes that make the task more suitable for examining

subsequent memory. This includes using trial-unique nameable objects, slower presentation durations, and fewer trials (see **Methods** for more details). Participants' memory for the trial-unique objects was then tested with free recall. Lag-CRP curves, and analyses of recall based on different "in the zone" and "out of the zone" event segments, allowed us to test whether and how attentional fluctuations shape the temporal organization of memory.

It is worth emphasizing that addressing our question requires a synthesis of disparate paradigms that are used in the sustained attention and memory recall fields. To foreshadow our results, we found that striking the right balance between the opposing needs of these tasks can be a challenge — a challenge we attempted to tackle across 4 Studies. Our approach and results highlight the difficulties of studying the effects of attentional lapses on recall organization, and thus a secondary aim of our project is to raise awareness of these challenges and offer strategies for future research that tackles similar questions.

1.2 Study 1

1.2.1 Methods

1.2.1.1 Design

1.2.1.1.1 Participants

We conducted an a priori power analysis using G* Power (Faul et al., 2007, 2009). Due to the lack of prior work involving spontaneous attentional fluctuations and temporal organization of recall, we calculated the number of participants required to both a) replicate typical properties of the temporal organization of recall, including main effects of, and the interaction between, absolute lags and direction (Kahana et al., 2002; Spillers & Unsworth, 2011; Palombo et al., 2019; Diamond & Levine, 2020); and b) observe interactions between lag-CRP properties (absolute lag or direction) and other independent variables (Palombo et al., 2019;

Diamond & Levine, 2020; this was done to approximate interactions between lag-CRP properties and attentional state). For 80% power and an alpha of 0.05 in a within-participant design, we determined that the minimum required sample size was 50 participants. We opted to exceed that to counteract effect size overestimation resulting from publication bias (Brand et al., 2008; Bakker et al., 2012). We therefore report data from 65 participants (M_{age} = 25.17 ± 6.67 years, $M_{education}$ = 14.78 ± 2.23 years; see Table 1 for demographics). We do not report data from an additional 15 participants, who were excluded due to image loading errors (N = 6), low response rate during the encoding task (3 SD from the group mean; N = 3), recall recording issues (N = 3), and incomplete data due to technical problems (N = 1).

Of the final sample, 22 participants were recruited from the Columbia University participant pool. They completed the study in the lab and were compensated with course credit. Because of the COVID-19 pandemic and the related disruption of research activities, the remaining participants (N = 43) were recruited through Prolific (www.prolific.co). They participated in an online version of the same experiment hosted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020).

Participants were 18 to 40 years of age, fluent in English, and resided in the US (inclusion criteria were specified in Prolific prior to recruitment). Both groups of participants provided informed consent in accordance with the Columbia University Institutional Review Board. No statistically significant differences were observed between the in-person and Prolific samples in any measure of interest (all ps > 0.14 for all main effects and interactions involving the 'sample' variable); thus, all data analyses include the combined sample. Nevertheless, for completeness, we report statistics to compare the two groups for effects of interest.

	Study 1	Study 2	Study 3	Study 4
Sample size	65	68	68	124
Age **(M ± SD)**	25.17 ± 6.67	22.62 ± 5.15	20.09 ± 2.20	21.42 ± 6.26
Education **(M ± SD)**	14.78 ± 2.23	13.99 ± 1.65	14.03 ± 1.46	13.63 ± 1.54
Recruitment **Method**	22 CU 43 Prolific	48 CU 20 Prolific	68 CU	119 CU 5 Prolific
Gender	33 F 28 M 3 NB 1 DNS	44 F 23 M 1 NB	33 F 35 M	84 F 37 M, 2 NB 1 DNS
Race	33 W 16 A 8 B/AA 5 BR 1 AI/AN 1 ME 1 O	39 W 15 A 10 B/AA 4 BR	35 W 19 A 7 B/AA 4 BR 1 AI/AN 2 O	67 W 30 A 14 B/AA 4 BR 3 AI/AN 3 H/L 1 ME 2 DNS
Ethnicity	59 NH/L 6 H/L	60 NH/L 8 H/L	54 NH/L 14 H/L	98 NH/L 26 H/L

Table 1: Summary of participant demographics across studies. For age and education, we report the mean ± standard deviation. For recruitment method, CU = Columbia University participant pool. For gender, F = female, M = male, NB = non-binary, DNS = did not specify. For race, W = white, A = Asian, B/AA = Black or African American, BR = bi-racial, AI/AN = American Indian or Alaskan Native, ME = Middle Eastern, O = other. For ethnicity, NH/L = not Hispanic or Latino, and H/L = Hispanic or Latino. In Study 4, three participants wrote in Latino for race. determined that the minimum required sample size was 50 participants. We opted to exceed that to counteract effect size overestimation resulting from publication bias (Brand et al., 2008; Bakker et al., 2012). We therefore report data from 65 participants (Mean age = 25.17 ± 6.67 years, Mean education = 14.78 ± 2.23 years; see Table 1 for demographics). We do not report data from an additional 15 participants, who were excluded due to image loading errors (N = 6), low response rate during the encoding task (<80%, N = 2), outlier response accuracy during the encoding task (>3 SD from the group mean; N = 3), recall recording issues (N = 3), and incomplete data due to technical problems (N = 1).

1.2.1.1.2 Stimuli

We chose 191 images of objects from pre-curated object databases such as SOLID (Frank et al., 2020), stimuli from the Mnemonic Similarity Task (Yassa & Stark, 2011 ; https://faculty.sites.uci.edu/starklab/mnemonic-similarity-task-mst/), Interaction Envelope (Bainbridge & Oliva, 2015a, 2015b; http://www.wilmabainbridge.com/datasets.html), and the Bank of Standardized Stimuli (Brodeur et al., 2014; https://sites.google.com/site/bosstimuli/). Color images were converted to grayscale.

These images were assigned to 5 study blocks using the OptSeg tool (Siegelman, 2019; https://github.com/msieg/OptSeg_Reproducible), which pseudo-randomizes the stimuli into lists while controlling for semantic similarity between constituent words. Semantic similarity between two items was measured as the cosine distance between the 300-dimensional GloVe vectors of the object names (Siegelman, 2019). Given a pool of words, the algorithm constructs lists so that items within a list are as semantically dissimilar as they can be. This allowed us to focus on temporal organization within lists that minimized opportunities for semantic organization given the stimuli available (Manning & Kahana, 2012). All 191 images were provided to the algorithm to have some leeway in assigning semantically matched and optimized lists. We created 5 lists of 30 images each using this procedure. The 5 stimulus lists were then randomized to 5 study blocks with within-block randomization of image order. The remaining 41 images were used for the practice block before the main task.

Although this approach attempts to minimize semantic similarity within a list, each list still had categorical structure. Specifically, within a list, items typically belonged to one of these categories: Musical Instruments, Animals, Plants, Food and Drinks, Tools, Home and Furniture,

or Sports and Games. Thus, we could not rule out semantic clustering in our task, a topic we return to in the **General Discussion**.

1.2.1.1.3 Procedure

The experiment was composed of 5 blocks, each of which included a study phase, a distractor phase, and a recall phase (Figure 1).

In each of the 5 study phases, participants viewed a series of 30 trial-unique images of common objects. The objects transitioned from one into another over 5 seconds. Each object remained on the screen for 1 second before the start of the next transition. Every 0.5 seconds the first object's opacity decreased by 10% while the succeeding object's opacity increased by 10%. This slow transition ensures that there is no capture of attention by abrupt image onsets, and helps induce more fluctuations of attention (Esterman et al., 2013; Rosenberg et al., 2013). For each presented image, participants were asked to judge if the depicted object was "smaller or larger than a shoebox" using one of two keys. These two keys were counterbalanced across participants. Participants could respond at any time once the object had started fading in (Figure 1).

A distractor phase followed each study phase, to introduce a delay between memory encoding and subsequent recall. Participants were presented with a sequence of 2-digit additions and subtractions. One solution to each question was provided and participants had to judge if the solution was correct or incorrect using one of two keys. Each trial began with a fixation cross presented for 2 seconds, followed by the arithmetic question for 5 seconds. 10 arithmetic questions were presented in each phase.

Memory recall immediately followed this distractor task. During each recall phase, participants were asked to verbally recall, in any order, as many objects as they could from the

preceding study phase. This phase was recorded using a microphone. There was a time limit of 2-2.5 minutes (see below), and a countdown timer was displayed to indicate the time remaining.

Before beginning the experimental blocks, participants performed a practice block. The practice block was identical to the experimental blocks except that only 15 objects were presented in the study phase.

Both versions of the task (in-person and online) were similar except for two minor modifications. The verbal recording for the in-person version was 2.5 minutes long for each recall portion; for the online version, it was 2 minutes long (the maximum allowed on Gorilla). Because participants rarely used the entire 2.5 minutes for recall in the in-person version, this was a minor change that had no measured impact on performance (overall recall was not significantly different between the in-person and online groups ($t_{48.18}$ = 1.32, p = 0.19, Cohen's ds = 0.35, 95% CI [-2.53, 12.16]) nor was recall organization different between the groups, all ps > 0.14 for interactions involving sample [online vs. in-person] in lag-CRP and event segment analyses discussed below). Additionally, the practice items for the in-person version were chosen from the leftover images after the Optseg assignment, and different participants were shown a different set of images. However, for the online version all participants saw the same 15 images in a randomized order.

We used 30 items per block — more than many standard list-learning recall studies (see Sederberg et al., 2010 for a few examples) because we expected that longer blocks (i.e., longer lists of items) would be necessary to reliably induce attentional fluctuations. Indeed, much longer block lengths (400 trials or more) are used in studies of sustained attention (Esterman et al., 2013; Rosenberg et al., 2013; Decker et al., 2020; Wakeland-Hart et al., 2022). We, therefore,
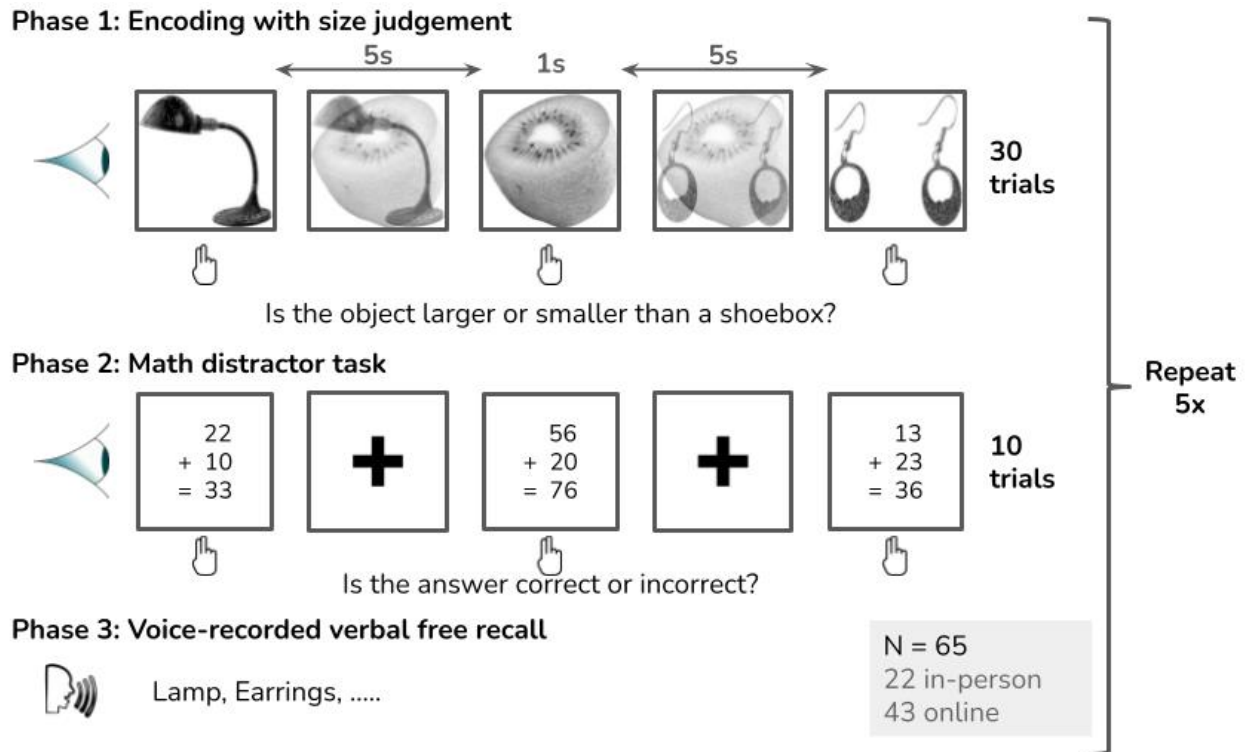
Figure 1: Schematic of the experimental design for Study 1. Participants were presented with a series of grayscale objects in the study phase. They were asked to judge if each object is smaller or larger than a shoebox. Object images gradually transitioned from one into another. Following a math distractor task, participants were asked to verbally recall objects from the study phase, in any order that they choose. This 3-phase sequence was repeated in 5 blocks, with 30 images encoded each time. Study 2 was similar except that there were 3 blocks of 80 items each and the task was to press a button for non-food items and withhold a response for food items. Study 3 had 3 blocks of 60 items each and the task was to press a button for color images and withhold a response for grayscale images. In Study 3, there were no gradual transitions between items; each image was shown for 3s with a 2s inter-item interval. Study 4 was similar to Study 2, except that there were 2 blocks of 120 images each, and the trial duration was reduced from 6s to 4s.

had fewer lists than typical list-learning recall studies, because more items were encoded per list.

In Studies 2-4, we further varied the number of lists and the number of items within them, in the

hope of striking a good balance between obtaining good recall (which is typically better with

shorter lists) and inducing attentional fluctuations (which typically occurs with long blocks).

1.2.1.2 Analyses

1.2.1.2.1 Defining attentional states at encoding

Attentional states were defined using response times (RTs) for the judgements made during the study phase. To assign responses to trials, we used a similar procedure as that used by Esterman et al. (2013) and Rosenberg et al. (2013). One modification was made to accommodate the slower image transitions in our tasks (3s or 5s in our Studies rather than 800ms in prior work). Furthermore, in our Studies, images stayed 100% coherent for 1s (rather than no "pause" at 100% coherence or a 400ms "pause" at 100% coherence in prior work). Because of these changes, there was a large peak in the RT distributions when an image was 90-100% coherent, and the majority of responses (~83-93%, across our Studies) came when an image was fading in between 50-100% coherence. There were few responses when an image was fading out from 100% coherence because of the relatively long 1s "pause" at the 100% coherent image (during which many responses were made). We therefore adjusted the image coherence thresholds used in prior work to reflect our observed RT distributions.

RTs were therefore assigned as follows. RTs were calculated relative to the beginning of each image transition (i.e., when it was starting to fade in). Responses when an image was fading in between 50% and 100% coherence were assigned to that image; this reflected the majority of trials. On the rarer trials for which a response occurred when an image was fading in between 0-50% (i.e., the previous image was fading out from <100% to 50%), an algorithm was applied as follows (following Esterman et al., 2013 and Rosenberg et al., 2013). First, the algorithm assigned unambiguous correct responses — that is, responses to the "go" category that occurred between 50% and 100% coherence of the image that was fading in. Second, the remaining, ambiguous responses (i.e., those from 0-50% coherence of the image fading in or multiple

25

responses) were assigned to an adjacent trial if one of the two had no response. If both adjacent

trials had no response, ambiguous responses were assigned to the closest trial, unless one was a

"no-go" trial, in which case participants were given the benefit of the doubt that they had

correctly omitted. Finally, if there were multiple responses that could be assigned on any one

trial, the fastest response was selected. The variance time course analyses described below were

also run with an alternative approach that simply assigned a response to the image that was

currently fading in; the pattern of results was largely unaffected with this alternative procedure.

A variance time course (VTC) analysis was performed on the RT data using the

procedure in Esterman et al., 2013. This procedure enables the identification of two attentional

states ("in the zone" and "out of the zone") based on the variability of RT. Each correct trial was

assigned a value corresponding to the absolute deviation of the trial RT from the mean within-

block RT. Trials without a response and trials with an incorrect response were not included in

this step. Next, this value was linearly interpolated (from the neighboring two trials) for trials

without a response and trials with an incorrect response. If only one trial was available for

interpolation (i.e., because trials at the beginning or end of a block do not have two surrounding

trials), then RT was not interpolated and such trials were not assigned an attentional state. Then,

following the procedure used by Esterman et al., (2013) and Rosenberg et al., (2013), a Gaussian

smoothing kernel was applied. We used the smth function in the smoother R package

(https://CRAN.R-project.org/package=smoother), and set the window to 4 trials (full-width at

half-maximum) and method to Gaussian. This procedure therefore integrated information from

the 4 surrounding trials (approximately 24 seconds, as in Rosenberg et al., 2013). Finally, a

median split was performed on these smoothed variance time course values, dividing the trials

into those with lower RT variability (i.e., RTs closer to the mean; "in the zone" states) and higher

RT variability (i.e., RTs farther away from the mean; "out of the zone" states).

This VTC analysis therefore assumes that RTs that are too fast or too slow (with respect

to their distance from the mean RT) reflect a poor attentional state. The rationale for this is that

RTs that are too fast reflect habitual responding or being on "autopilot" and RTs that are too

slow reflect disengagement from, or inefficient processing of, the task at hand (Esterman et al.,

2013; Kane & McVay, 2012; Yamashita et al., 2021). Indeed, this approach has been successful

at predicting attentional lapses (e.g. Esterman et al., 2013; Rosenberg et al., 2013; Esterman et

al., 2014; Rosenberg et al., 2015; Madore et al., 2020; also see Yamashita et al., 2021). However,

it is critical to note that there are other approaches to characterizing attentional states based on

RT. For example, some studies of sustained attention have shown that faster (vs. slower) RTs are

linked to more online errors and worse subsequent recognition memory (e.g., Robertson et al.,

1997; Cheyne et al., 2006, 2009; Kane & McVay, 2012; deBettencourt et al., 2018; Wakeland-

Hart et al., 2022). Yet, counter to these approaches, other studies link longer eye fixations and

slower response times to mind wandering and attentional lapses (e.g., Smallwood et al., 2003;

Weissman et al., 2006; Feng et al., 2013; Yanko & Spalek, 2013; Kam & Handy, 2014;

Henríquez et al., 2016; Krasich et al., 2018; Zhang et al., 2021). Thus, slower (vs. faster) RTs

have sometimes been linked to better attentional states and sometimes to worse attentional states,

and the VTC analysis considers RTs that are either too fast or too slow (relative to the mean) to

reflect a bad attentional state.

To accommodate this diversity in identifying better vs. worse attentional states from RTs,

we provide a secondary analysis for each study in which we examined attentional states defined

by faster RTs vs. slower RTs. Although the VTC approach was our a priori analysis of choice,

27

this allowed us to examine whether alternative definitions of attentional states lead to different results with respect to online errors and subsequent memory.

1.2.1.2.2 Errors at encoding

In prior studies, "out of the zone" (vs. "in the zone") attentional states were associated with more errors in sustained attention tasks (Esterman et al., 2013; Rosenberg et al., 2013). To determine if we could replicate those findings, we examined whether there was a difference between the two attentional states in the number of errors made during the encoding task, which was designed to be similar to the sustained attention tasks used in prior studies. First, objects depicted by each image were classified as being either larger than a shoebox (e.g., helicopter, treadmill), smaller than a shoebox (e.g., onion, key), or ambiguous (e.g., cowboy hat, soda bottle). Next, errors were calculated as the number of incorrect responses made to the objects that were unambiguously classified (i.e., responses to ambiguous objects were never counted as incorrect). (Note that accuracy was therefore the percentage of presented items with a correct response, for which responses to ambiguous objects were always counted as correct). Finally, we examined whether the number of errors made during "out of the zone" attentional states was higher than the number of errors made during "in the zone" states. Group-level analyses were conducted with a paired-samples t-test.

1.2.1.2.3 Recall performance

We calculated recall performance as the percentage of items that were correctly recalled across all blocks. To examine whether recall performance differed by attentional state, we calculated recall performance for each attentional state ("in the zone" or "out of the zone") as the percentage of items that were encoded in that particular attentional state that were correctly recalled. Note that because these attentional states are defined by a median split of the encoding

RTs, the same number of items are encoded in each state. Group-level analyses were conducted with a paired-samples t-test.

1.2.1.2.4 Temporal organization of recall

We measured the temporal organization of recall using lag-Conditional Response Probability (lag-CRP) curves (Kahana, 1996). The lag-CRP curve measures the probability of recalling two items successively as a function of their relative position, or lag, at encoding. To plot these curves, we first obtain the encoding lag between all pairs of successively recalled items, where the lag is the difference between their serial positions at encoding. The lag can be positive or negative, depending on whether the subsequent item recalled was encoded after (positive lag) or before (negative lag) the preceding item. The observed number of recall transitions at each lag is then divided by the number of opportunities to make a recall transition at that lag, e.g., all the times a participant could have recalled an item at a +1 lag, regardless of whether or not they did (Kahana, 1996). This yields the probabilities plotted in the lag-CRP curve, i.e., actual transitions divided by possible transitions at each lag. Repetitions and intrusions are masked from this analysis: transitions to and from repetitions (recalled items that had also been recalled earlier) and intrusions (items recalled from a prior study list) are excluded (Kahana, 1996).

Lag-CRP curves depict two characteristic features of the temporal organization of recall: forward asymmetry and temporal contiguity. Forward asymmetry refers to the higher likelihood of recalling in the forward vs. backward direction (i.e., higher conditional probability of recall for positive lags vs. negative lags). Temporal contiguity refers to the higher probability of recalling items together if they were encoded closer together in time (Kahana, 1996; Healey et

al., 2019). This is seen as a peak in the lag-CRP curves: recall is more likely for items at ±1 lag, and recall probability decreases gradually with increasing lags.

Our primary hypothesis was that "in the zone", vs. "out of the zone", attentional states are more conducive to maintaining a temporal context representation. If so, there should be a difference in the temporal organization of recall between the attentional states: temporal contiguity and/or forward asymmetry should be stronger for "in the zone" vs. "out of the zone" attentional states. To test this, we constructed separate lag-CRP curves for the two attentional states. First, individual pairs of successively recalled items were labelled as being in the same state (i.e., both encoded during an "in the zone" state or both encoded during an "out of the zone" state) or being a transition pair from one state to another. To calculate the lag-CRP curves separately for each state, we only considered pairs that were in the same state; transition pairs were excluded from analysis (see **Recall Transitions by Event Segment** for consideration of state transitions). Actual transitions were calculated as the lag between two successively recalled items based on their encoding position. These actual transitions were divided by the number of times a transition of a given lag could have possibly occurred irrespective of attentional states (i.e., all possible transitions) to yield the conditional response probability. The conditional response probability for each lag was calculated within-block for each attentional state. These values were then averaged across blocks, resulting in one CRP value at each lag for each attentional state for each participant.

To test for a statistical difference in the temporal organization of recall between the two attentional states, we performed a three-way repeated-measures ANOVA with attentional state ("in the zone" vs. "out of the zone"), absolute lag (1 to 29), and recall direction (forward vs. backward) as factors. We tested the sphericity assumption using Mauchly's test of sphericity and

applied the Greenhouse-Geisser correction when this assumption was not satisfied. This analysis also allowed us to test for typical properties of lag-CRP curves, by looking for main effects of absolute lag and direction, and the interaction between them.

Separate lag-CRP curves for each attentional state (and the associated 3-way ANOVA) required us to discard some data (recall transitions *between* "in the zone" and "out of the zone" states and participants who did not have any successive recall transitions within "in the zone" states or within "out of the zone" states). Because of this, we also show the overall lag-CRP curves for each Study (using all the data, irrespective of attentional state at encoding), and report analyses of the overall lag-CRP curve with a two-way (direction x lag) ANOVA in **Table 2**. Finally, because there is a possibility that including all lags could mask differences in forward asymmetry — which is typically most prominent at nearby lags — we conducted an additional analysis. We examined whether there was a difference between the two attentional states in forward asymmetry at the closest lags of ± 1, by performing a two-way repeated-measures ANOVA with lag (+1 vs. -1) and attentional state ("in the zone" vs. "out of the zone") as factors (Diamond & Levine, 2020).

The approach we take in the above analysis — dividing actual transitions within an attentional state by all possible transitions at a given lag *regardless of attentional state at encoding* — effectively asks: of all possible recall transitions available to the participant at lag of X, how many were for items encoded "in the zone" vs. "out of the zone"? For example, we divide the number of +1 transitions between "in the zone" items by the total number of possible +1 transitions regardless of attentional state at encoding; because the denominator is therefore the same for the two attentional states, we will refer to this as the "same denominator" approach.

An alternative analysis choice would be to divide actual transitions by the number of times a transition of a given lag could have possibly occurred *within that same attentional state*; e.g., divide +1 transitions for "in the zone" items by the number of times a +1 transition *to other "in the zone" items* could have occurred. We will refer to this as the "state-specific denominator" approach.

In simulated data, we found that these approaches sometimes yield identical *differences* between attentional states, although the *absolute value* of the CRP will necessarily be higher when the denominator is specific to a given attentional state. In other cases, the "state-specific denominator" approach underestimated differences in temporal structure between conditions, while in yet other cases, it was more accurate than the "same denominator" approach — particularly when recall rates differed between conditions but temporal structure did not. Because this "state-specific denominator" approach can be argued to be valid in some situations, we repeated all the analyses (across all Studies) with this alternative approach — but the pattern of results did not change (**Table S1**).

1.2.1.2.5 Recall Transitions by Event Segment

The above temporal organization analyses do not differentiate between qualitatively different types of transitions that are possible within each attentional state. Because individuals fluctuate between "in the zone" and "out of the zone" periods during encoding, each instance of an attentional state can be considered its own "event segment" (**Figure 2**). These different segments are ignored in the lag-CRP analysis above, which simply considers each attentional state as a whole. Furthermore, our "in the zone" and "out of the zone" lag-CRPs ignored recall transitions *between* "in the zone" and "out of the zone" states. Yet, considering these types of transitions is important: research on event boundaries has shown that successive recall of

adjacently encoded items is more likely when those items are encoded in the same vs. different

event (Ezzyat & Davachi, 2010; DuBrow & Davachi, 2013, 2016; Heusser et al., 2018; also see

DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014) and recall may "leap" between cognitively

similar but temporally distant events (Chan et al., 2017). We therefore designed an analysis to

test whether fluctuations between attentional states can act as event boundaries that reproduce

the phenomena observed in research on events.

This analysis allowed us to test whether 1) recall transitions within an "event segment"

are more likely for "in the zone" vs. "out of the zone" attentional states; and 2) whether recall for

"in the zone" states is more likely to "leap" from one event segment to another, bypassing items

that were encoded in an "out of the zone" attentional state (Chan et al., 2017; also see Heusser et

al., 2018). To examine these hypotheses, we considered all pairs of successively recalled items,

and calculated the number of transitions made during recall for each type of transition noted in

Figure 2 (same state, same event; same state, different event; different state, different event),

separately for each attentional state ("in the zone" vs. "out of the zone"). We divided the number

of actual transitions in each bin by the number of opportunities to transition to another item that

falls within the same transition type regardless of attentional state at encoding. The resulting

value is therefore the conditional probability of each transition type. For example, we divided the

number of "same state, same event" transitions for "in the zone" items by the number of

opportunities to make "same state, same event" transitions to other items, regardless of

attentional state at encoding (i.e., the total number of possible "same state, same event"

transitions between two "in the zone" items and between two "out of the zone" items). This

analysis therefore asks: of all possible event transitions of a specific type (e.g., "same state, same

event"), how many were for items encoded "in the zone" vs. "out of the zone"? This is analogous to the "same denominator" approach we used for the lag-CRPs.
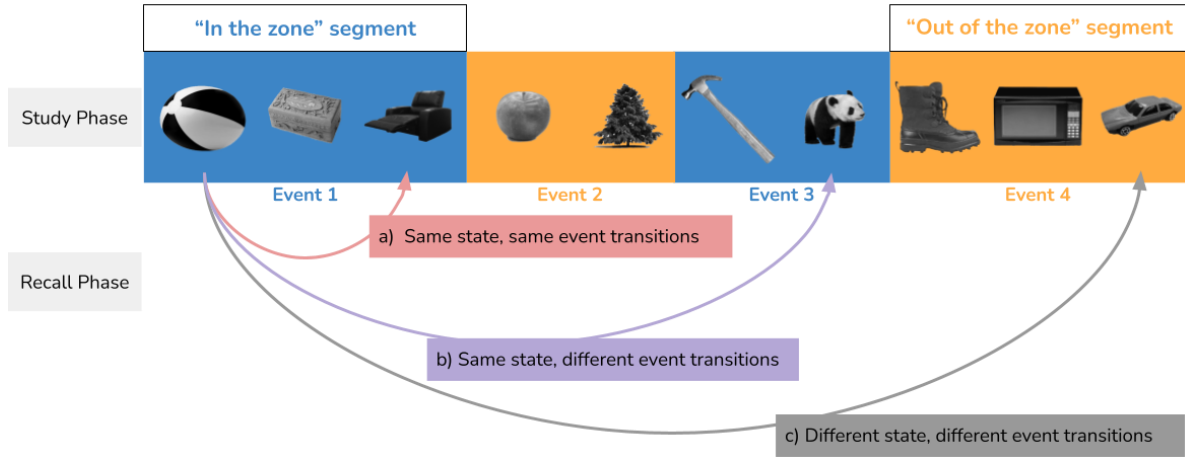
Figure 2. Types of transitions at recall. Individuals fluctuate between "in the zone" (blue) and "out of the zone" (orange) attentional states during encoding. Each instance of each attentional state can be considered its own "event segment". For example, the ball and the recliner are encoded in the same attentional state and the same event segment within that state (a: "same state, same event"), whereas the ball and the panda are encoded in the same attentional state but during different event segments (b: "same state, different event"). The ball and the car are encoded in different attentional states and therefore, by necessity, different event segments as well (c: "different state, different event"). In the recall phase, given the recall of one object, the transition to the second object can be therefore categorized as one of three types: 1) same state, same event; 2) same state, different event; and 3) different state, different event.

For completeness, we also re-ran this analysis using the "state-specific" denominator approach, in which we divided the number of transitions in each bin by the number of opportunities to transition to another item that falls within the same type of transition *for that attentional state*. For example, we divided the number of "same state, same event" transitions for "in the zone" items by the number of opportunities to make "same state, same event" transitions to items encoded "in the zone".

In both the "same denominator" and "state-specific denominator" approaches, these conditional probabilities were calculated separately for each block, and then averaged across

34

blocks for each participant. We then performed a two-way repeated-measures ANOVA with the type of transition (3 levels) and attentional state (2 levels) as factors. The pattern of results was the same for the "same denominator" and "state-specific" denominator approaches; here, we report the results for the "same denominator" approach to be consistent with the lag-CRP analysis. The results for the "state-specific" denominator approach can be found in **Table S1**.

1.2.1.2.6 Bayes Factors

We computed Bayes Factors using the BayesFactor package in R (https://cran.r-project.org/web/packages/BayesFactor/index.html; Morey & Rouder, 2011; Rouder & Morey, 2011; Rouder et al., 2012, 2013). Bayes Factors quantify the strength of evidence provided by the data for one hypothesis vs. another.

We report Bayes Factors (BF) in terms of support for the null hypothesis, such that values greater than 3 indicate evidence for the null while values less than 1/3 indicate evidence for the alternative hypothesis (Kass & Raftery, 1995; Jeffreys, 1961; Wagenmakers et al., 2011). Specifically, Bayes Factors between 1/3-3 indicate no evidence for either the null or alternative hypothesis; Bayes Factors between 3-20 or between 1/20-1/3 provide evidence for the null or alternative hypothesis, respectively; Bayes Factors between 20-150 or between 1/150-1/20 provide strong evidence for the null or alternative hypothesis, respectively; and Bayes Factors greater than 150 or less than 1/150 indicate very strong evidence in favor of the null or alternative hypothesis, respectively. For the sake of completeness and transparency, we report Bayes Factors for all of our main findings (in all studies and in **Table 3**). However, we note that care must be taken when computing and interpreting Bayes Factors because they can be subject to many limitations (e.g., sensitivity to prior selection), including limitations similar to those of p-values (e.g., binary treatment of a continuous measure). For discussions about the limitations

of Bayes Factors, see Liu & Aitkin, 2008; Simonsohn, 2014; Simmons et al., 2016; Gelman & Carlin, 2017.

Bayes Factors for paired t-tests were computed with the function *ttestBF()* with *paired* set to *TRUE*, and other arguments left as the default; the inverse was then obtained to provide the Bayes Factor in terms of support for the null. For main effects and interactions from ANOVAs, we computed Bayes Factors using the function *lmBF()* with default priors and 500,000 iterations. Bayes Factors for main effects were computed by dividing the Bayes Factor for the model excluding the main effect (and all interactions with the effect of interest) by the Bayes Factor for the model including the main effect, but no interactions with the effect of interest. Similarly, to compute the Bayes Factor for an interaction effect, we divided the Bayes Factor obtained from a model excluding the interaction effect (but including all other main effects and interactions) by the Bayes Factor for a model including the interaction of interest (Rouder et al., 2012).

In addition to Bayes Factors for the effects of interest, we also included Bayesian model comparisons for the lag-CRP and event segment analyses. Here, we compared a null model that did not include attentional state as a factor (whether as a main effect or interaction term) to a full model that included attentional state and its interactions (Liu & Aitkin, 2008; Kruschke, 2011). The Bayes Factor for the null model was then divided by the Bayes Factor for the full model (both computed using the function lmBF) to obtain evidence for the null (i.e., values larger than 3 indicate evidence for the null model).

We also provide traditional null-hypothesis significance testing throughout the paper (i.e., determining whether results are statistically significant or not based on whether p-values are less than 0.05). In almost all cases, p-values and Bayes Factors are in agreement. There are, however, some situations in which p-values are less than 0.05 but evidence from Bayes Factors does not

strongly support the alternative hypothesis; this only happens in rare cases and does not affect our main conclusions. In all analyses testing our main hypotheses — with respect to the temporal organization of recall based on attentional state at encoding — interpretation of p-values and Bayes Factors leads to the same conclusion.

1.2.2 Results

1.2.2.1 Defining attentional states at encoding

In the encoding task, participants viewed images of objects and judged each as being larger or smaller than a shoebox. Overall, mean response time (RT; defined from the onset of an image fading in; see **Methods** and **Figure 1**) was 3.94 (SD = 0.58s). Median RT was 4.00s.

We defined attentional states by performing a variance time course (VTC) analysis on the encoding phase RTs (Esterman et al., 2013). This procedure identifies two attentional states: the good "in the zone" attentional state (trials with lower RT variability, i.e., RTs closer to the mean) and the worse "out of the zone" attentional state (trials with higher RT variability, i.e., RTs farther away from the mean). **Figure 3A** shows the VTC analysis for one sample participant.

The mean length of a continuous "in the zone" segment was 3.59 trials (SD = 0.46) and the mean length of a continuous "out of the zone" segment was 3.59 trials (SD = 0.53; Note that each trial was 6 seconds long). The mean number of fluctuations within a block (i.e., the number of times participants transitioned from one state to another) was 7.38 (SD = 1.04). The number of trials within a continuous segment ranged from 1 to 14, for both "in the zone" and "out of the zone" states, across all blocks and participants.

As an initial validation check of the VTC approach, we examined whether RT variability changed over the course of the experiment, as would be predicted if participants gradually lost focus or became fatigued over time. Indeed, RTs became progressively more variable across

blocks ($F_{3.49,223.21} = 4.91$, p = 0.0015, $\eta_p^2 = 0.07$), consistent with the idea that more variable RTs may index worsening attention and differentiate "in the zone" from "out of the zone" states. These attentional states were used to examine accuracy on the encoding task and subsequent recall performance, described below.

1.2.2.2 More encoding errors during "out of the zone" attentional states

Participants generally performed very well on the encoding task ("Is this object smaller or larger than a shoebox?"). The mean percentage of trials participants responded to was 97.82% (SD = 3.11%, Median = 99.33%). Mean accuracy (as defined in **Methods**) was 88.66% (SD = 7.12%, Median = 91.33%).

We next examined errors in the encoding task as a function of attentional state. Prior studies have shown that "out of the zone", vs. "in the zone", attentional states are associated with more errors (Esterman et al., 2013; Rosenberg et al., 2013). We replicated these findings in the current study. A paired samples t-test revealed that participants made significantly more errors in the encoding task during an "out of the zone" attentional state (mean $\pm$ SD: 10.31 $\pm$ 6.67) compared to an "in the zone" attentional state (7.15 $\pm$ 5.32; $t_{64} = 4.87$, p < 0.0001, Cohen's dz = 0.60, 95% CI [1.86, 4.45], BF < 1/150, **Figure 3B**). Thus, the VTC analysis is successful in identifying fluctuations between better and worse attentional states, even in our modified procedure.
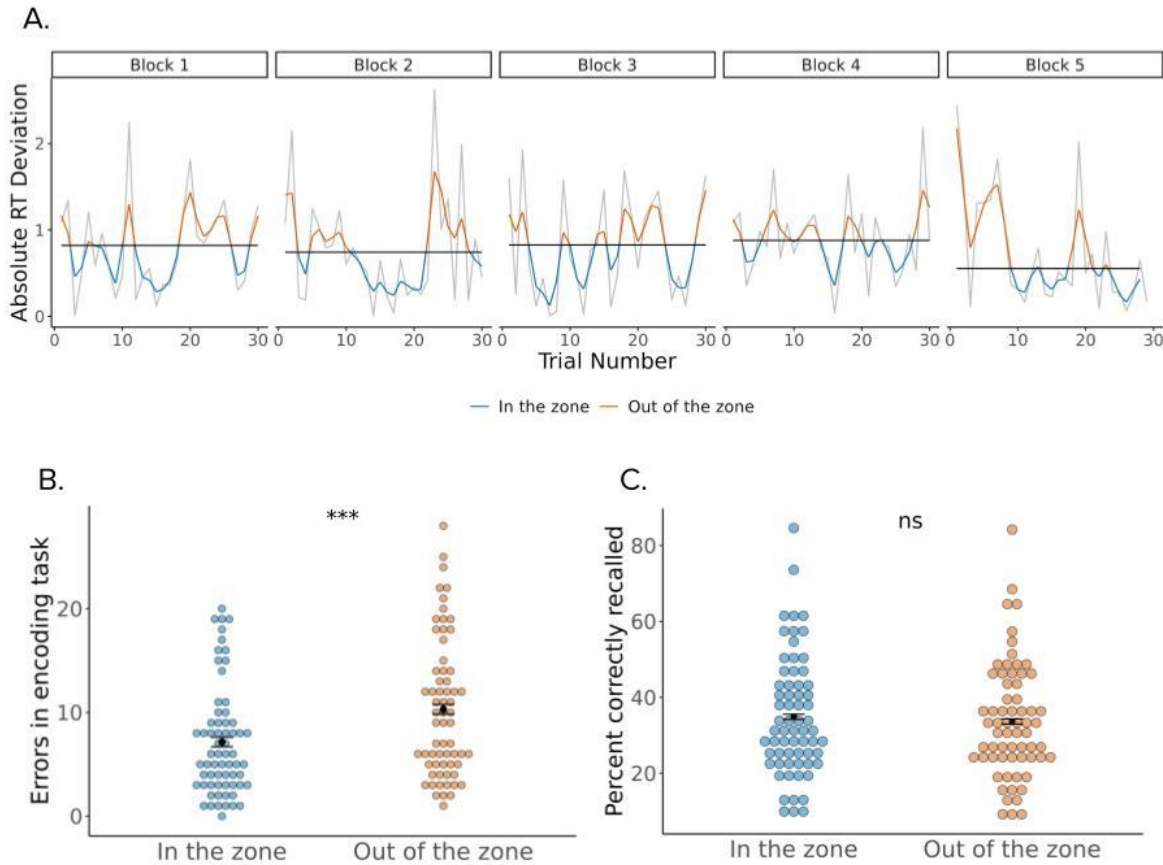
Figure 3. Encoding task performance and recall performance in Study 1. Encoding errors differ between attentional states but recall performance does not. A. Variance Time Course (VTC) analysis for a sample participant. Two attentional states were identified by a median split of smoothed, absolute RT deviations from the mean: 1) an "in the zone" attentional state (blue) with lower RT variability, i.e., RTs closer to the mean and, 2) an "out of the zone" attentional state (orange) with higher RT variability, i.e., RTs farther away from the mean. Horizontal black lines indicate the median absolute RT deviation per block. Gray curves indicate raw (unsmoothed) RT deviation per block. B. Individual points show the number of encoding judgement errors made by each participant during "in the zone" and "out of the zone" attentional states. Participants made significantly more encoding errors during the "out of the zone" state. C. Individual points show the percentage of items correctly recalled by each participant as a function of whether items were encoded "in the zone" or "out of the zone". There was no difference in recall performance between the two states. Black points in panels B & C indicate the mean of the measure; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". *** $p < .0001$. ns = not statistically significant.

As an additional validation check of whether RTs are sensitive to attentional fluctuations in our task, we compared RTs before vs. after an error, as done in prior work (e.g., Dutilh et al., 2012; Fortenbaugh et al., 2015). To that end, we used a linear mixed-effects model to predict RT from trial position (pre vs. post error), with participant-level random effects for the intercept (following the procedure used in Decker et al., 2020). We found reliable post-error speeding: RTs after an error were significantly faster than those before an error (β = -230.82, S.E. = 74.15, t_{1955} = -3.11, p = 0.0019, 95% CI [-376.15, -85.49]). Although counterintuitive, several other studies have found post-error speeding (Notebaert et al., 2009; Purcell & Kiani, 2016; Williams et al., 2016; Damaso et al., 2020); this may occur if participants occasionally "zone out" or mind wander, leading to inefficient processing of an item on a given trial and subsequent speeding once participants catch themselves being off task. Thus, this analysis provides additional evidence that RTs can index attention in our task.

1.2.2.3 Overall recall does not differ between the two attentional states

We next turned to examining memory for the objects viewed during the encoding task. Mean recall accuracy, calculated as the percentage of items correctly recalled across all blocks, was 34.28% (SD = 14.68).

We then separately examined recall performance based on whether items were encoded "in the zone" or "out of the zone". We hypothesized that recall performance would be superior for "in the zone" attentional states. However, we did not find a significant difference in recall performance between "in the zone" (Mean ± SD: 34.86% ± 15.70) and "out of the zone" (33.64% ± 15.20) attentional states (t_{64} = 1.26, p = 0.21, Cohen's dz = 0.16, 95% CI [-0.72, 3.17], BF = 3.46, **Figure 3C**). Thus, even though these attentional states differed in *online* task performance, subsequent recall was surprisingly not different.

	Study 1	Study 2	Study 3	Study 4
Lag (1 to 29)	Nearby > far away $F_{12.42, 794.89} = 11.28$ $p < 0.0001$ $\eta_p^2 = 0.15$	Nearby > far away $F_{11.36, 761.14} = 14.47$ $p < 0.0001$ $\eta_p^2 = 0.18$	Nearby > far away $F_{11.53, 772.73} = 11.49$ $p < 0.0001$ $\eta_p^2 = 0.15$	Nearby > far away $F_{13.46, 1655.35} = 7.34$ $p < 0.0001$ $\eta_p^2 = 0.06$
Direction (Forward vs. Backward)	$F_{1,64} = 0.33$ $p = 0.57$ $\eta_p^2 = 0.005$	Forward > backward $F_{1,67} = 15.77$ $p = 0.0002$ $\eta_p^2 = 0.19$	$F_{1,67} = 0.002$ $p = 0.97$ $\eta_p^2 = 0.00003$	$F_{1,123} = 1.35$ $p = 0.25$ $\eta_p^2 = 0.01$
Lag x Direction	$F_{14.79, 946.73} = 1.45$ $p = 0.12$ $\eta_p^2 = 0.02$	Forward > backward for nearby > far away $F_{15.85,1062.11} = 1.88$ $p = 0.019$ $\eta_p^2 = 0.027$	Forward > backward for nearby > far away $F_{15.86,1062.47} = 1.84$ $p = 0.023$ $\eta_p^2 = 0.027$	$F_{16.06,1975.52} = 1.27$ $p = 0.21$ $\eta_p^2 = 0.01$
+1 vs. -1 transitions	$t_{64} = 1.89$ $p = 0.063$ Cohen's dz = 0.23 95% CI [-0.002,0.059]	$t_{67} = 3.59$ $p = 0.0006$ Cohen's dz = 0.44 95% CI [0.01,0.04]	$t_{67} = 3.32$ $p = 0.002$ Cohen's dz = 0.40 95% CI [0.01,0.04]	$t_{123} = 1.92$ $p = 0.06$ Cohen's dz = 0.17 95% CI [-0.0003,0.018]

Table 2. Summary of two-way ANOVA (direction x absolute lag) results for the overall lag-CRP curves in each Study. Green shading indicates statistically significant results ($p < 0.05$) and gray shading indicates results that are not statistically significant.

1.2.2.4 No differences in temporal contiguity or forward asymmetry between the two attentional states

Although overall recall was not different between "in the zone" and "out of the zone" attentional states, it is possible that there may be subtle differences in *how* information is recalled. We therefore turned to our main analyses of interest, which explore the temporal organization of recall. We hypothesized that "in the zone" attentional states are more conducive to maintaining temporal context representations that facilitate temporally organized recall. We

therefore used lag-CRP curves to test whether temporal contiguity and/or forward asymmetry were stronger for "in the zone" vs. "out of the zone" states.

Figure 4A shows the overall lag-CRP curve, across participants and blocks, regardless of attentional state at encoding (see **Table 2** for analyses of the overall lag-CRP curve); **Figure 4B** shows the lag-CRP curves, across participants and blocks, separately for items that were encoded "in the zone" and "out of the zone". These curves depict the probability of recalling two items successively based on their relative position, or lag, at encoding. Note that, across all of our Studies, the CRP values at nearby lags are lower than what is typically observed in many recall studies (e.g., Kahana, 1996; Sederberg et al., 2010; Healey & Kahana, 2014). This is likely because the lists used in the current set of Studies are longer than those of most list-learning recall studies and, concomitantly, more recall transitions are available for each item. Thus, recall transitions could be spread across more lags. This is consistent with the finding that lag-CRP values at nearby lags tend to be lower for longer lists (Healey et al., 2019).

We conducted a three-way repeated-measures ANOVA on the lag-CRP measures (i.e., lag-conditional recall shown in **Figure 4B**) with attentional state ("in the zone" vs. "out of the zone"), absolute lag (1 to 29), and direction (forward vs. backward) as factors (results summarized in **Table 3**). We expected that we might find 1) an interaction between attentional state and absolute lag, indicating that nearby recall transitions would be more likely for items encoded "in the zone" vs. "out of the zone" and 2) an interaction between attentional state and direction, indicating a stronger forward asymmetry bias for items encoded "in the zone". We found a significant main effect of absolute lag ($F_{9.07,\,580.52} = 9.60$, $p < 0.0001$, $\eta_p^2 = 0.13$, BF < 1/150): during recall, individuals were more likely to transition to items that were encoded nearby vs. farther away. We did not find a significant main effect of direction ($F_{1,64} = 0.074$, $p = $

		Study 1	Study 2	Study 3	Study 4
	Online errors (In vs. Out)	Out > In $p < 0.0001$ BF < 1/150	Out > In $p < 0.0001$ BF < 1/150	$p = 0.92$ BF = 7.47	Out > In $p < 0.0001$ BF < 1/150
	Overall recall (In vs. Out)	$p = 0.21$ BF = 3.46	$p = 0.56$ BF = 6.35	$p = 0.14$ BF = 2.56	In > Out $p = 0.03$ BF = 0.96
lag-CRP	Lag (1 to 29)	Nearby > far away $p < 0.0001$ BF < 1/150	Nearby > far away $p < 0.0001$ BF < 1/150	Nearby > far away $p < 0.0001$ BF < 1/150	Nearby > far away $p < 0.0001$ BF < 1/150
	Direction (Forward vs. Backward)	$p = 0.79$ BF = 37.41	Forward > backward $p = 0.02$ BF = 1.96	$p = 0.80$ BF = 38.88	$p = 0.20$ BF = 24.95
	Attentional State (In vs. Out)	$p = 0.74$ BF = 37.92	$p = 0.57$ BF = 31.22	$p = 0.20$ BF = 12.39	In > Out $p = 0.0485$ BF = 2.08
	Lag x Direction	$p = 0.33$ BF > 150	Forward > backward for nearby > far away $p = 0.003$ BF = 0.56	Forward > backward for nearby > far away $p = 0.007$ BF = 18.22	Forward > backward for nearby > far away $p = 0.02$ BF > 150
	Lag x Attentional State	$p = 0.64$ BF > 150	$p = 0.39$ BF > 150	$p = 0.66$ BF > 150	$p = 0.84$ BF > 150
	Direction x Attentional State	$p = 0.33$ BF = 20.99	$p = 0.79$ BF = 26.88	$p = 0.73$ BF = 26.99	$p = 0.47$ BF = 29.46
	Lag x Direction x Attentional State	$p = 0.29$ BF > 150	$p = 0.19$ BF > 150	$p = 0.51$ BF > 150	$p = 0.90$ BF > 150
lag-CRP +/- 1	Lag (+1 vs. -1)	+1 > -1 $p = 0.043$ BF = 0.52	+1 > -1 $p = 0.0009$ BF = 0.024	+1 > -1 $p = 0.0007$ BF = 0.028	+1 > -1 $p = 0.002$ BF = 0.043
	Attentional State (In vs. Out)	$p = 0.22$ BF = 4.25	$p = 0.66$ BF = 6.89	$p = 0.46$ BF = 6.36	$p = 0.59$ BF = 8.96
	Lag x Attentional State	$p = 0.76$ BF = 5.09	$p = 0.26$ BF = 2.86	$p = 0.26$ BF = 3.08	$p = 0.92$ BF = 7.20
Event Analysis	Transition Type	$p < 0.0001$ BF < 1/150	$p < 0.0001$ BF < 1/150	$p < 0.0001$ BF < 1/150	$p < 0.0001$ BF < 1/150
	Attentional State (In vs. Out)	$p = 0.85$ BF = 8.87	$p = 0.54$ BF = 7.77	$p = 0.13$ BF = 4.41	$p = 0.27$ BF = 6.93
	Transition Type x Attentional State	$p = 0.83$ BF = 19.13	$p = 0.53$ BF = 15.08	$p = 0.26$ BF = 10.66	$p = 0.58$ BF = 25.54

Table 3. Summary of results across Studies 1-4. Green shading indicates statistically significant results ($p < 0.05$) and gray shading indicates results that are not statistically significant. The interpretation of Bayes Factors (BF) is described in the main text. F-values, t-values, degrees of freedom, effect sizes, and confidence intervals are reported in the main text. In = "in the zone" attentional state; Out = "out of the zone" attentional state, defined with the VTC analysis.
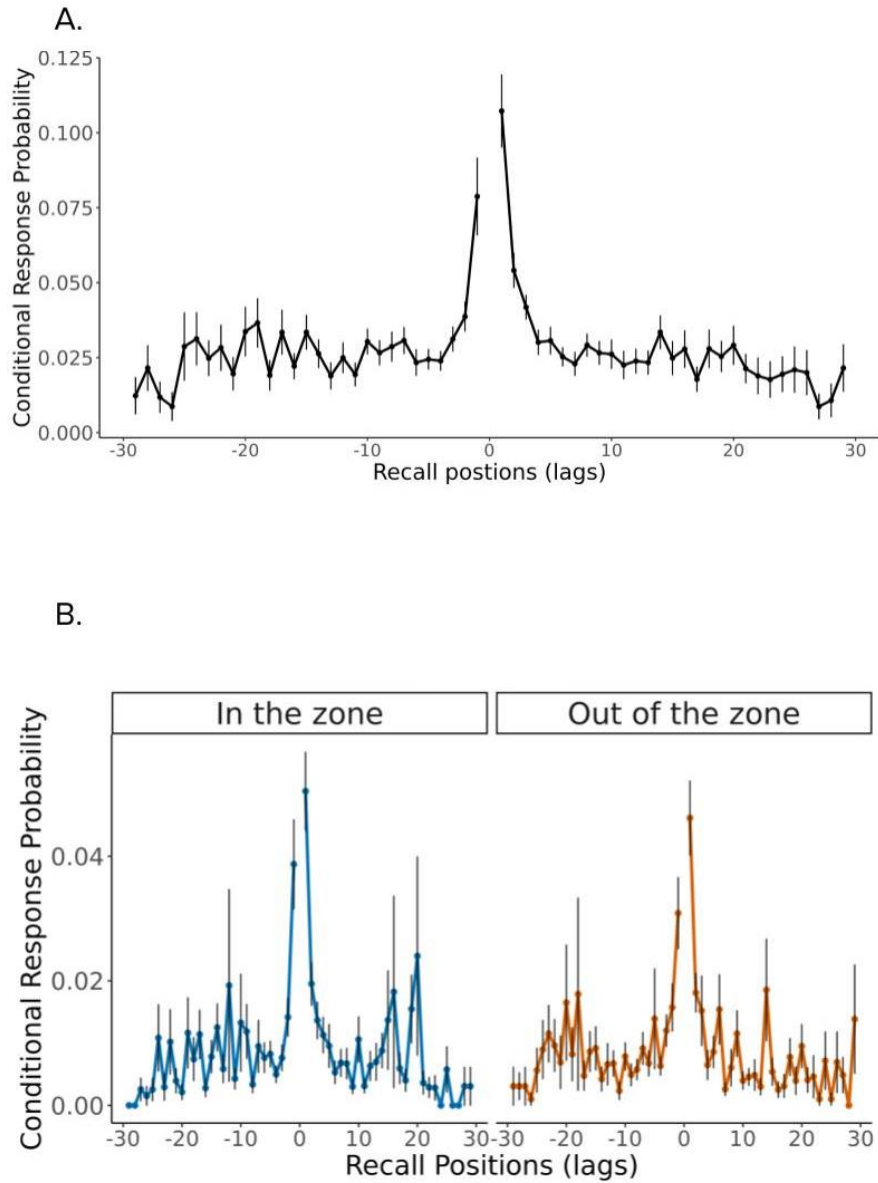
Figure 4. Lag-CRP curves overall and by attentional state for Study 1. A. Overall lag-CRP curve across participants and blocks. B. Lag-CRP curves plotted separately for items encoded "in the zone" (left) and "out of the zone" (right). There was no difference between the two attentional states in the temporal contiguity or forward asymmetry of recall. Error bars represent the standard error.

0.79, $\eta_p^2 = 0.001$, BF = 37.41) nor an interaction between direction and absolute lag ($F_{8.44, 540.05} = 1.15$, $p = 0.33$, $\eta_p^2 = 0.02$, BF > 150).

We next examined main effects and interactions involving attentional state. There was no main effect of attentional state ($F_{1,64} = 0.11$, p = 0.74, $\eta_p{}^2 = 0.002$, BF = 37.92). Additionally, there was no interaction between attentional state and direction ($F_{1,64} = 0.96$, p = 0.33, $\eta_p{}^2 = 0.015$, BF = 20.99), no interaction between attentional state and absolute lag ($F_{7.66\ 490.37} = 0.75$, p = 0.64, $\eta_p{}^2 = 0.012$, BF > 150), nor a three-way interaction between absolute lag, direction, and attentional state ($F_{8.27,\ 529.45} = 1.21$, p = 0.29, $\eta_p{}^2 = 0.02$, BF > 150). Hence, there was no significant difference between the attentional states in temporal contiguity or forward asymmetry.

We next examined forward asymmetry differences between the two attentional states at the nearby lags of $\pm$ 1. This was done to determine whether the lack of a forward asymmetry difference between the attentional states was due to inclusion of all lags in the CRP curves: forward symmetry is sometimes most pronounced for nearby lags. From a two-way repeated-measures ANOVA with lag (+1 vs. -1) and attentional state ("in the zone" vs. "out of the zone") as factors, we did not find an effect of attentional state ($F_{1,64} = 1.51$, p = 0.22, $\eta_p{}^2 = 0.023$, BF = 4.25) nor an interaction between attentional state and lag ($F_{1.62} = 0.091$, p = 0.76, $\eta_p{}^2 = 0.001$, BF = 5.09).

Finally, we conducted a Bayesian model comparison in which we compared the three-way ANOVA reported above (lag x direction x attentional state) to a null model without an attentional state variable (lag x direction). This analysis revealed very strong evidence in favor of the null model (BF > 150).

All together, across these lag-CRP analyses, we did not find any differences in the temporal organization of recall for items encoded "in the zone" vs. "out of the zone".

1.2.2.5 No differences in event transition types between the two attentional states

The lag-CRP analyses above examined temporal organization differences between the two attentional states, but they do not take into account qualitatively different types of transitions that could occur within an attentional state (such as transitions between different event segments) or transitions from one state to another (see **Figure 2**). However, it is possible that attentional fluctuations act in a similar way to event boundaries (Ezzyat & Davachi, 2010; DuBrow & Davachi, 2013, 2016; Heusser et al., 2018), such that recall is more temporally clustered within segments than across segments, and may occasionally "leap" between segments of a similar cognitive state (Chan et al., 2017).

We therefore examined recall as a function of the type of transition (**Figure 2**). If "in the zone" (vs. "out of the zone") attentional states are more conducive to maintaining a temporal context representation, and this temporal context representation is reinstated every time an individual is "in the zone", two predictions could be made. First, that recall transitions within an "event segment" may be more likely for items encoded "in the zone" vs. "out of the zone" (i.e., same state, same event transitions; **Figure 2**), and second, that recall "leaps" to a different event segment in the same attentional state may be more likely for items encoded "in the zone" vs. "out of the zone" (i.e., same state, different event transitions).

To test this, for each type of transition ("same state, same event"; "same state, different event"; "different state, different event"), we calculated the number of transitions made during recall divided by the number of opportunities to make such transitions (see **Methods: Recall Transitions by Event Segment**). We performed a two-way repeated-measures ANOVA with transition type (3 levels) and attentional state ("in the zone" vs. "out of the zone") as factors. We expected to find an interaction between attentional state and transition type. Specifically, we

expected to find more "same state, same event" and "same state, different event" transitions for "in the zone" vs. "out of the zone" attentional states.

We found a main effect of transition type ($F_{1.07, 68.51} = 32.24$, $p < 0.0001$, $\eta_p^2 = 0.33$, BF < 1/150). There was no main effect of attentional state ($F_{1,64} = 0.037$, $p = 0.85$, $\eta_p^2 = 0.0006$, BF = 8.87) nor a significant interaction between attentional state and transition type ($F_{1.28, 81.66} = 0.086$, $p = 0.83$, $\eta_p^2 = 0.001$, BF = 19.13).
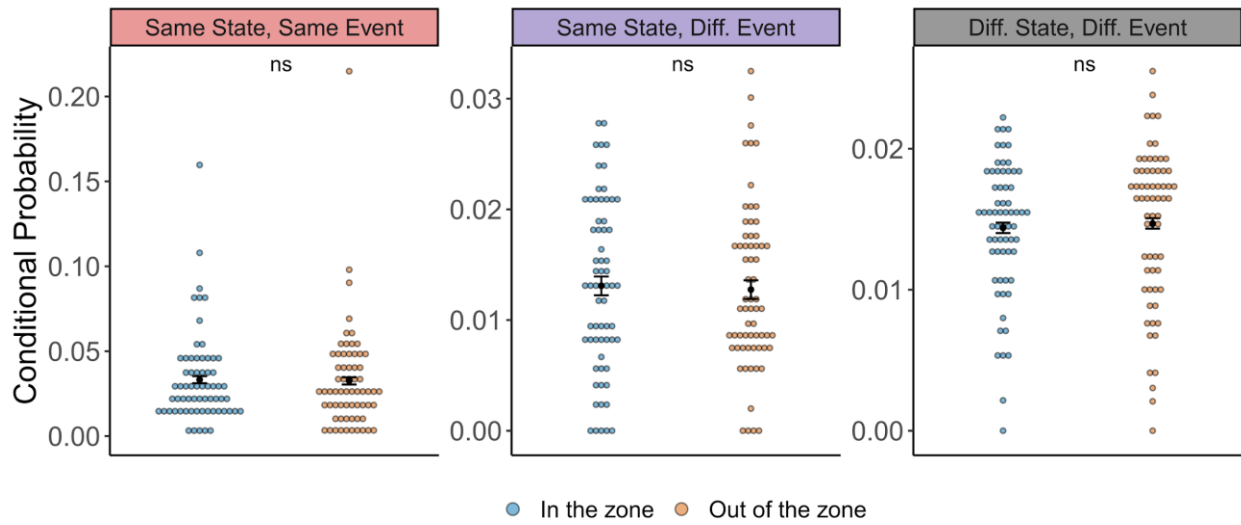
Figure 5. Recall transitions as a function of event type for Study 1. Recall transitions are shown based on whether they occurred within an "event segment" of a particular attentional state (same state, same event), across event segments of a given attentional state (same state, different event), or between attentional states (different state, different event; see Figure 2). There were no significant differences between the two attentional states in any transition type. Individual points indicate the conditional probability of each transition type (i.e., the number of times each transition type occurred divided by the number of opportunities to make a transition of that type) for each individual, separately for items encoded "in the zone" and "out of the zone". Black dots indicate the mean; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". ns = not statistically significant.

Given the main effect of transition type, we conducted follow-up t-tests to compare them, collapsing across the two attentional states. We found that "same state, same event" transitions were significantly more likely than "same state, different event" transitions ($t_{64} = 5.77$, $p <$

0.0001, Cohen's dz = 0.72, 95% CI [0.013, 0.027]) and significantly more likely than "different

state, different event" transitions (t_{64} = 5.71, p < 0.0001, Cohen's dz = 0.71, 95% CI [0.012,

0.025]). "Different state, different event" transitions were also more likely compared to "same

state, different event" transitions (t_{64} = 2.13, p = 0.04, Cohen's dz = 0.26, 95% CI [0.00010,

0.003]).

This pattern of results (**Figure 5**) is consistent with the temporal contiguity effect:

participants are more likely to make recall transitions to nearby items. In particular, items

encoded in the same state and same event are the closest to one another ("same state, same

event" transitions; e.g., recall transitions within Event 1 in **Figure 2**), and are more likely to be

recalled together than items that span events ("same state, different event" and "different state,

different event" transitions in **Figure 2**).

As for the lag-CRP analysis, we conducted a Bayesian model comparison in which we

compared the two-way ANOVA reported above (transition type x attentional state) to a null

model without an attentional state variable (one-way ANOVA with transition type). This

analysis revealed very strong evidence in favor of the null model (BF > 150). Therefore, across

these analyses, we found no evidence that recall transitions across different event types varied

based on attentional state at encoding.

1.2.2.6 Secondary Analysis: Faster vs. Slower RTs

In addition to defining attentional states based on the VTC approach (in which RTs that

are either too fast or too slow are considered to reflect a poor attentional state), we conducted a

secondary analysis in which we divided attentional states into those characterized by faster vs.

slower RTs, in line with other research that makes this distinction (e.g., deBettencourt et al.,

2018; Decker et al., 2020; Wakeland-Hart et al., 2022). We therefore analyzed the data the same

way as we did for the VTC analysis (including, for example, interpolating RTs on incorrect trials), except that trials were divided by a median split into those associated with faster vs. slower RTs (rather than RTs closer vs. further from the mean).

There were no differences in online errors during attentional states characterized by faster vs. slower RTs ($t_{64} = 1.90$, p = 0.06, Cohen's dz = 0.24, 95% CI [-0.05, 2.05], BF = 1.36). Note that the numerical direction of this effect (more errors for attentional states associated with slower vs. faster RTs) is the opposite from that expected based on some similar prior work, in which more errors were made following faster vs. slower RTs (deBettencourt et al., 2018; Wakeland-Hart et al., 2022). It is, however, consistent with findings that longer eye fixations and slower response times are associated with mind wandering and attentional lapses (e.g., Smallwood et al., 2003; Weissman et al., 2006; Feng et al., 2013; Yanko & Spalek, 2013; Kam & Handy, 2014; Henríquez et al., 2016; Krasich et al., 2018; Zhang et al., 2021).

Subsequent recall was not different between attentional states associated with faster vs. slower RTs at encoding ($t_{64} = 1.21$, p = 0.23, Cohen's dz = 0.16, 95% CI [-0.77, 3.15], BF = 3.66). The lag-CRP analysis revealed only a main effect of absolute lag (nearby > far away; $F_{12.68, 811.72} = 7.96$, p < 0.0001, $\eta_p^2 = 0.11$, BF < 1/150); all other main effects and interactions were not statistically significant (all *p*s > 0.26; all other BFs > 20.40).

The event segment analysis revealed a main effect of transition type ($F_{1.04, 66.24} = 35.41$, p < 0.0001, $\eta_p^2 = 0.36$, BF < 1/150). Follow-up t-tests showed significantly more "same state, same event" transitions compared to "same state, different event" transitions ($t_{64} = 5.90$, p < 0.0001, Cohen's dz = 0.73, 95% CI [0.01, 0.03]), and "different state, different event" transitions ($t_{64} = 6.10$, p < 0.0001, Cohen's dz = 0.76, 95% CI [0.015, 0.029]). "Different state, different event" transitions were also more likely than "same state, different event" transitions ($t_{64} = 2.03$,

$p = 0.047$, Cohen's $dz = 0.25$, 95% CI [0.00002, 0.003]). There was no main effect of attentional state ($F_{1,64} = 0.25$, $p = 0.62$, $\eta_p^2 = 0.004$, BF = 8.50) nor an attentional state by transition type interaction ($F_{1.12, 71.44} = 1.24$, $p = 0.28$, $\eta_p^2 = 0.02$, BF = 10.35).

A Bayesian model comparison for the lag-CRP analysis, which compared the above model to a null model without an attentional state variable, revealed very strong evidence in favor of the null model (BF > 150). The same approach for the event segment analysis revealed strong evidence in favor of the null model (BF = 86.32).

Thus, this alternative characterization of attentional states also failed to reveal differences in the temporal organization of recall. This alternative approach is, however, more difficult to interpret than the VTC analysis: unlike the VTC analysis, which was successful in detecting differences in online performance for "in the zone" vs. "out of the zone" attentional states, this alternative approach did not reliably predict online errors. Thus, this alternative approach may not be as powerful in identifying fluctuating attentional states.

1.2.3 Discussion

We hypothesized that "in the zone" attentional states, vs. "out of the zone" states, are more conducive to maintaining temporal context representations that can facilitate temporally organized recall. We found no support for this hypothesis. There was no difference between attentional states in overall recall; there was also no difference in forward asymmetry or temporal contiguity as assessed by the lag-CRP curves; and finally, there was no difference in the types of event transitions made in recall.

We replicated prior results in showing more online errors (in the encoding task) during "out of the zone" vs. "in the zone" attentional states. Yet, subsequent recall was not different overall between the two attentional states, and no measures of recall organization showed a

difference. One possibility is that our task only yielded moderate attentional fluctuations, which were not strong enough to produce effects on recall. This may be because our task deviated from the original gradCPT in several ways (Esterman et al., 2013; Rosenberg et al., 2013). The original gradCPT requires a habitual response to frequent trials and a withheld response to infrequent trials. Here, we had participants provide a binary judgement (using one of two keys) on each trial, for which the responses were similar in frequency. Thus, the traditional manipulation might be more effective in inducing "out of the zone" states because 1) the same response is made most of the time and 2) the judgement used in the current task might be more difficult and subjective, and thus may have required more focused attention.

Furthermore, the subjectivity of the judgements used in this Study may have made it difficult to separate "in the zone" and "out of the zone" states. This may have occurred if RTs to ambiguous images (e.g., cowboy hat, soda bottle) were slow, as participants considered how either response (bigger or smaller than a shoebox) could be justified. If that was the case, responses to ambiguous items may be incorrectly considered "out of the zone", obscuring true differences between attentional states. We conducted two additional analyses to test this possibility. First, we examined if RTs to ambiguous items were slower than those to unambiguous items; that was indeed the case ($t_{64} = 3.88$, $p = 0.0003$, Cohen's dz = 0.48, 95% CI [-0.34, -0.11]). We next repeated all of our analyses after interpolating RTs for ambiguous items, using the same procedure used for error trials. The pattern of results was unchanged: the statistically significant results were in online errors ("out of the zone" > "in the zone"; $t_{64} = 3.85$, $p = 0.0003$, Cohen's dz= 0.48, 95% CI [1.20, 3.79]); the main effect of absolute lag in the lag-CRP analysis (nearby transitions > far away transitions; $F_{9.87, 631.61} = 8.49$, $p < 0.0001$, $\eta_p^2 = 0.12$; all other ps > 0.22); and the main effect of transition type in the event segment analysis (same

51

pattern as reported in the main analysis; $F_{1.08,68.92} = 29.59$, $p < 0.0001$, $\eta_p^2 = 0.32$; all other $ps > 0.45$). Thus, inclusion of ambiguous items had no detectable effect on our analyses of interest.

A final difference between our Study and prior work is that our blocks were relatively short: 3 minutes relative to 8 minutes in a traditional gradCPT. These short blocks may not have induced strong enough attentional fluctuations to see large effects on recall organization. In Study 2, we modified our task to address these limitations.

1.3 Study 2

1.3.1 Overview

In Study 2, we sought to address limitations of Study 1 that may have made attentional fluctuations relatively weak and thus, limited our chance of seeing strong effects on subsequent recall. First, we changed our task to be aligned with the traditional gradCPT by using a go/no-go approach. Most trials were "go" trials in which a participant made a response to non-food items. On a minority of trials ("no-go"), which occurred 10% of the time, a food item was presented and participants had to withhold their response. This approach should make the "go" response habitual, making it more likely that individuals will "zone out" due to the repetitive nature of the task. We did not include any stimuli for which a "food" vs. "not food" judgment would be ambiguous (e.g., animals); this addresses a limitation of Study 1, in which some ambiguous images were used. Second, we made our encoding blocks longer (3 blocks of 8 minutes, rather than 5 blocks of 3 minutes in Study 1). This was done with the hope of encouraging stronger periods of "zoning out". To that end, we increased our block length by adding more items per block: 80 items per block rather than the 30 in Study 1. A small pilot study indicated that participants recalled fewer words with the longer blocks; we therefore added a block beyond what was needed to (roughly) balance the total number of items in Study 1 and Study 2 (i.e., 3

blocks, rather than 2, of 80 items each). This allowed us to match the average number of words

recalled across Study 1 and Study 2.

1.3.2 Methods

1.3.2.1 Design

1.3.2.1.1 Participants

We report data from 68 participants (M_{age} = 22.62 ± 5.15, $M_{education}$ = 13.99 ± 1.65; see

Table 1 for demographics). We do not report data from an additional 15 participants, who were

excluded due to image loading errors (N = 3), low response rate during the encoding task (<80%,

N = 7), outlier response accuracy during the encoding task (>3 SD from the group mean; N = 1),

and recall recording issues (N = 4). Of the final sample, 48 participants were recruited from the

Columbia University participant pool and the rest (20 participants) were recruited through

Prolific (www.prolific.co). All participants completed an online version of the task hosted on the

Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020). Informed consent was obtained in

accordance with the Columbia University Institutional Review Board.

1.3.2.1.2 Stimuli

Stimuli were identical to Study 1 with the following exceptions. We chose 240 images

instead of 191 from the pre-curated databases. Color images were converted to grayscale. 90% of

the images (216 images) were non-food (i.e., inedible) items and 10% (24 images) were food

items. No ambiguous stimuli were included (e.g., animals). The 240 images were divided into 3

lists of 80 images each (8 food, 72 non-food images). For this and subsequent Studies, the

stimulus lists were created by manually ensuring there were an equal number of items from a

category (e.g., tools, furniture) in each list. This manual sorting was used instead of the OptSeg

algorithm from Study 1 because there were fewer lists, and it was thus more tractable to do the

balancing manually (OptSeg was relatively slow to run and required manual checking of the generated lists; thus, it was more efficient to make the lists manually).

1.3.2.1.3 Procedure

The procedure was identical to Study 1 with the following exceptions. The experiment consisted of 3 blocks, each of which included a study phase, a distractor phase, and a recall phase (**Figure 1**). In each study phase, participants viewed 80 trial-unique items, which transitioned slowly from one into another as in Study 1. For each presented image, participants were asked to judge if the depicted object was "a food or non-food item". Importantly, they were asked to press a button when it was a non-food item (the dominant category), but withhold their response when it was a food item. This change aligned our task with the traditional gradCPT, such that participants habitually pressed one response 90% of the time, which may make it more likely for them to "zone out".

The distractor phase was identical to that in Study 1. The recall phase was similar to Study 1. Participants were initially given 2 minutes to verbally free recall items from the study phase. Unlike Study 1, after the initial 2 minutes of recording (the maximum allowed on Gorilla), participants were given the option to begin recording for another 2 minutes if they wanted to recall more objects. This was done because the blocks in Study 2 were longer than those in Study 1; thus, we wanted to give participants more time for free recall if they needed it. Participants did not perform a practice block before beginning the task blocks, but were given video instructions on how to perform the task.

1.3.2.2 Analyses

1.3.2.2.1 Defining attentional states at encoding

Attentional states were defined using RTs for the judgements made by participants during the study phase, in a procedure similar to Study 1. Unlike Study 1, however, participants were supposed to withhold responses on some trials; thus, some correct responses did not have an associated RT. Therefore, we first calculated — for correct trials with a response — the absolute deviation of the trial RT from the within-block mean, as in Study 1. Next, RT deviations for trials without a response (whether correctly withheld on "no-go" food trials or incorrectly withheld on "go" non-food trials) and trials with an incorrect response were interpolated from the two surrounding trials, as done in other studies employing this method (Esterman et al., 2013; Rosenberg et al., 2013). All other steps were performed in an identical manner to Study 1. This resulted in trials being divided into two attentional states: "in the zone" states with lower RT variability (i.e., RTs closer to the mean) and "out of the zone" states with higher RT variability (i.e., RTs farther away from the mean; **Figure 6A**).

1.3.2.2.2 Errors at encoding

Similar to Study 1, we first sought to replicate the finding that "out of the zone" (vs. "in the zone") attentional states are associated with more errors (Esterman et al., 2013; Rosenberg et al., 2013). Errors in the encoding task were calculated as the sum of the number of incorrect button presses to a "no-go" food trial (commission errors) and the number of failures to respond to a "go" non-food trial (omission errors). We examined whether the number of errors made during "out of the zone" attentional states was higher than the number of errors made during "in the zone" states. Group-level analyses were conducted with a paired-samples t-test.

1.3.2.2.3 Recall performance

Recall performance and related analyses were identical to those in Study 1.

1.3.2.2.4 Temporal organization of recall

Analyses of temporal organization of recall were identical to those in Study 1. Note that in Study 1, actual and possible transitions ranged from -29 to +29 and the entire range was used in lag-CRP analyses. In Study 2, the range of actual and possible transitions is -79 to +79, because the length of the encoding list is 80 items. However, for the lag-CRP analyses of interest, we only used actual and possible transitions between -29 to +29. There were two reasons for this: First, transitions at the farther lags were rare (26 trials or fewer, across all 3 blocks for all 68 participants combined, for a given lag further away than $\pm$ 29), and hence the lag-CRP estimates were particularly noisy at those lags. Second, to facilitate comparison across studies, we opted to keep our analyses consistent by using the range from Study 1.

1.3.2.2.5 Recall Transitions by Event Segment

Analyses of recall transitions by event segment were identical to those in Study 1.

1.3.3 Results

1.3.3.1 Defining attentional states at encoding

In the encoding task, participants viewed images of objects and judged each as being a non-food item (with a button press) or a food item (by withholding a response). Overall, mean RT (defined from the onset of an image fading in; see **Methods**) was 3.74s (SD = 0.72). Median RT was 3.77s.

As before, we defined "in the zone" and "out of the zone" attentional states by performing a variance time course (VTC) analysis on the encoding phase RTs (see **Methods**). **Figure 6A** shows the VTC analysis for one sample participant in Study 2.

The mean length of an "in the zone" segment was 3.95 trials (SD = 0.42) and the mean length of an "out of the zone" segment was 3.90 trials (SD = 0.43; Note that each trial was 6 seconds long). The mean number of fluctuations within a block (i.e., the number of times participants transitioned from one state to another) was 19.47 (SD = 2.25). The number of trials within a segment ranged from 1 to 27 for "in the zone" states and from 1 to 23 for "out of the zone" states, across all blocks and participants.

As for Study 1, we sought an initial validation of the VTC approach by testing whether RTs were increasingly variable over the experimental session. Once again, RTs became progressively more variable over blocks ($F_{1.77,118.34} = 5.66$, $p = 0.006$, $\eta_p^2 = 0.08$), as individuals presumably became more fatigued and less focused. This lends support to the use of RT variability to differentiate "in the zone" and "out of the zone" states.
These attentional states were used to examine accuracy on the encoding task and subsequent recall performance, described below.

1.3.3.2 More encoding errors during "out of the zone" attentional states

Participants performed very well on the encoding task ("Is this object food or a non-food item?"). They responded to 99.41% (SD = 2.57%; Median = 100%) of the "go" non-food trials, which required a response. Mean accuracy (including correct responses on "go" trials and withheld responses on "no-go" trials) was 96.87% (SD = 2.74%; Median = 97.92%).

We next examined errors in the encoding task as a function of attentional state. We replicated prior studies (Esterman et al., 2013; Rosenberg et al., 2013) and Study 1: A paired samples t-test revealed that participants made significantly more encoding errors during an "out of the zone" attentional state (mean ± SD: 4.28 ± 3.78) compared to an "in the zone" attentional state (2.28 ± 2.43; $t_{67} = 4.83$, $p < 0.0001$, Cohen's $d_z = 0.59$, 95% CI [1.17, 2.83], BF < 1/150,

Figure 6B). Thus, the VTC analysis remains successful in identifying fluctuations between better and worse attentional states.

As in Study 1, we performed an additional validation check of whether RTs are sensitive to attentional fluctuations in our task, by comparing RTs before vs. after an error. Unlike Study 1 (for which participants should respond on every trial), there were two types of potential errors in Study 2: omission errors, for which participants fail to make a button response when they should, and commission errors, for which participants make a button response when they should have withheld it. We therefore examined RTs before vs. after these error types.

Participants were faster to respond after vs. before an omission error (β = -1148.6, S.E. = 152.00, $t_{310.40}$ = -7.56, p < 0.0001, 95% CI [-1446.53, -850.60]). This is consistent with participants "zoning out" and failing to respond to an image, and subsequently speeding up once they realize that a trial was missed. Conversely, participants responded slower after vs. before a commission error (β = 392.81, S.E. = 122.39, $t_{330.94}$ = 3.21, p = 0.002, 95% CI [152.94, 632.70]). This is consistent with fast, habitual responses leading to errors, with post-error slowing once an erroneous response has been detected. Together, post-error slowing after commission errors and post-error speeding after omission errors show that RTs can index attentional fluctuations in our task.
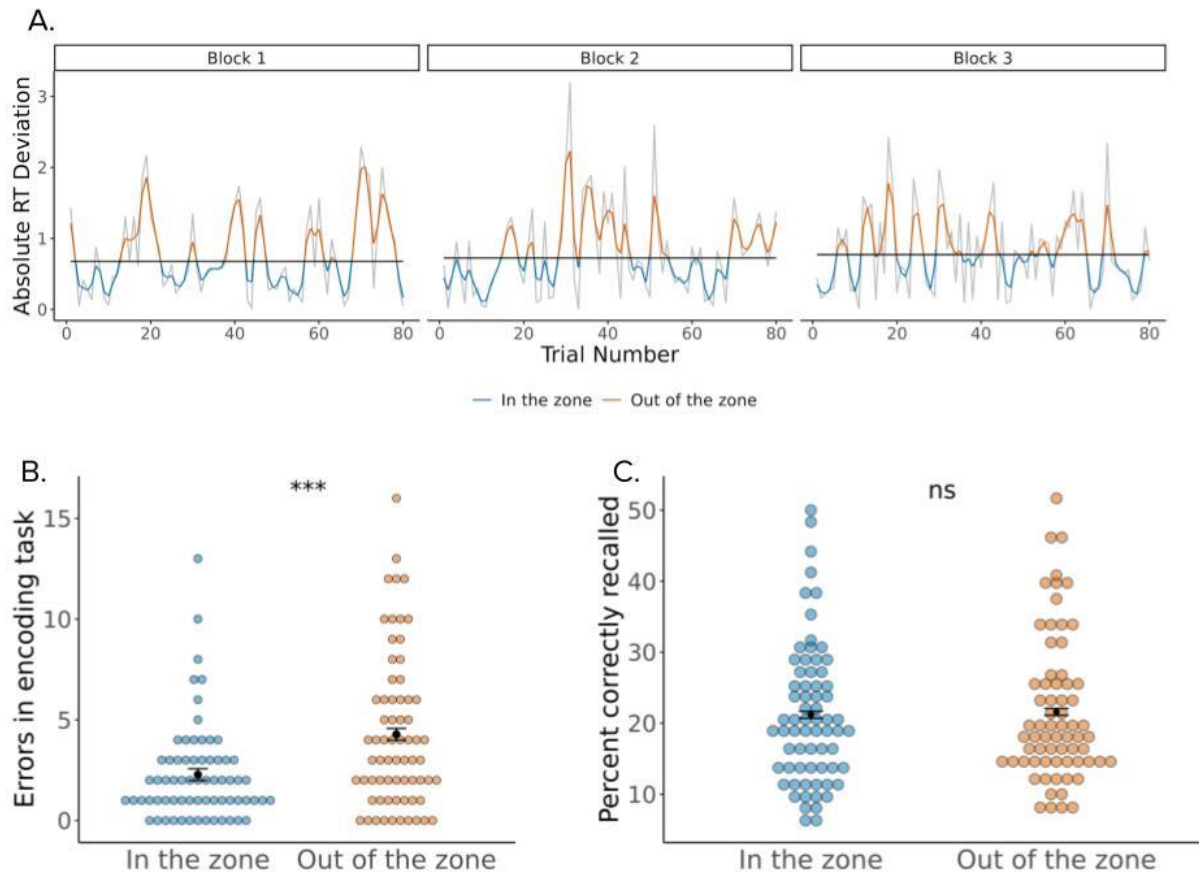
Figure 6. Encoding task performance and recall performance in Study 2. Encoding errors differ between attentional states but recall performance does not. A. Variance Time Course (VTC) analysis for a sample participant, depicting "in the zone" (blue) and "out of the zone" (orange) attentional states. Horizontal black lines indicate the median absolute RT deviation per block. Gray curves indicate raw (unsmoothed) RT deviation per block. B. Individual points show the number of encoding judgement errors made by each participant during "in the zone" and "out of the zone" attentional states. Participants made significantly more encoding errors during the "out of the zone" state. C. Individual points show the percentage of items correctly recalled by each participant as a function of whether items were encoded "in the zone" or "out of the zone". There was no difference in recall performance between the two states. Black points in panels B & C indicate the mean of the measure; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". *** p < .0001. ns = not statistically significant.

1.3.3.3 Overall recall does not differ between the two attentional states

We next turned to examining memory for the objects encoded during the study phase. Mean recall was 21.38% (SD = 9.71).

We then separately examined recall based on whether items were encoded "in the zone" or "out of the zone". As in Study 1, we did not find a significant difference in recall performance between "in the zone" (Mean ± SD: 21.19% ± 9.91) and "out of the zone" (21.60% ± 10.18) attentional states ($t_{67} = 0.59$, p = 0.56, Cohen's dz = 0.072, 95% CI [-0.97, 1.79], BF = 6.35, **Figure 6C**). Thus, as in Study 1, these states differed in performance during the encoding task but showed no differences in subsequent recall.

1.3.3.4 No differences in temporal contiguity or forward asymmetry between the two attentional states

As in Study 1, we constructed separate lag-CRP curves for "in the zone" vs. "out of the zone" attentional states based on successive recall of items encoded in the same state (**Figure 7B**; see **Figure 7A** and **Table 2** for overall lag-CRP). We conducted a three-way repeated-measures ANOVA with attentional state ("in the zone" vs. "out of the zone"), absolute lag (1 to 29), and direction (forward vs. backward) as factors (results summarized in **Table 3**).

We found a significant main effect of absolute lag ($F_{8.97, 600.97} = 16.22$, p < 0.0001, $\eta_p^2 = 0.20$, BF < 1/150): during recall, individuals were more likely to transition to items that were encoded nearby vs. farther away. We also found a significant main effect of direction ($F_{1,67} = 6.24$, p = 0.02, $\eta_p^2 = 0.09$, BF = 1.96) and an interaction between direction and absolute lag ($F_{12.47, 835.32} = 2.49$, p = 0.003, $\eta_p^2 = 0.04$, BF = 0.56). Thus, participants were more likely to recall items in the forward vs. backward direction, with this asymmetry being more pronounced for closer vs. farther lags.
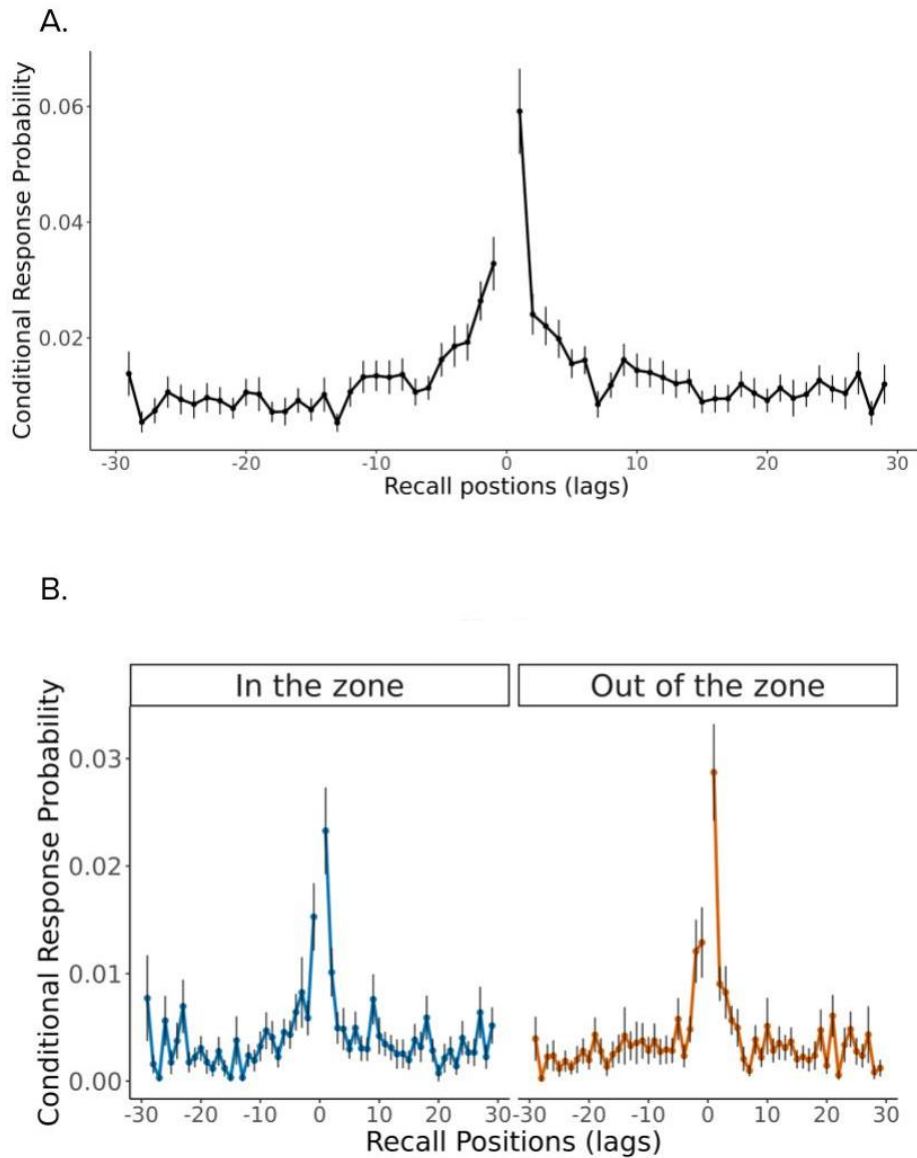
Figure 7. Lag-CRP curves overall and by attentional state for Study 2. A. Overall lag-CRP curve across participants and blocks. B. Lag-CRP curves plotted separately for items encoded "in the zone" (left) and "out of the zone" (right). There was no difference between the two attentional states in the temporal organization of recall (neither temporal contiguity nor forward asymmetry). Error bars represent the standard error.

We next examined main effects and interactions involving attentional state. There was no main effect of attentional state ($F_{1,67} = 0.32$, $p = 0.57$, $\eta_p^2 = 0.005$, BF = 31.22). There was also no interaction between attentional state and direction ($F_{1,67} = 0.08$, $p = 0.79$, $\eta_p^2 = 0.001$, BF =

26.88), no interaction between attentional state and absolute lag ($F_{12.37, 829.07} = 1.06$, p = 0.39, η_p^2 = 0.02, BF > 150), nor a three-way interaction between absolute lag, direction, and attentional state ($F_{13.11, 878.44} = 1.33$, p = 0.19, η_p^2 = 0.02, BF > 150). Hence, we did not see any differences in temporal contiguity or forward asymmetry bias of recall for items encoded "in the zone" vs. "out of the zone".

We conducted a follow-up analysis to compare the two attentional states at the lags of ± 1. This was done to examine whether including all lags in our repeated-measures ANOVA masked differences between the states that were more specific to close recall transitions. From a two-way repeated-measures ANOVA with lag (+1 vs. -1) and attentional state ("in the zone" vs. "out of the zone") as factors, we found only a significant main effect of lag ($F_{1,67} = 12.09$, p = 0.0009, η_p^2 = 0.15, BF = 0.024). The main effect of attentional state ($F_{1,67} = 0.20$, p = 0.66, η_p^2 = 0.003, BF = 6.89), and the interaction between attentional state and lag, were not statistically significant ($F_{1,67} = 1.28$, p = 0.26, η_p^2 = 0.02, BF = 2.86). The significant main effect of lag (+1 vs. -1) indicated that individuals are more likely to make forward vs. backward transitions at the closest lag. However, this forward asymmetry at the ± 1 lags was not different between the two states.

As for Study 1, we conducted a Bayesian model comparison in which we compared the three-way ANOVA reported above to a null model without an attentional state variable. As before, this analysis revealed very strong evidence in favor of the null model (BF > 150).

All together, across these analyses, we replicated the finding that recall is temporally organized. However, this temporal organization was not different between the two attentional states.
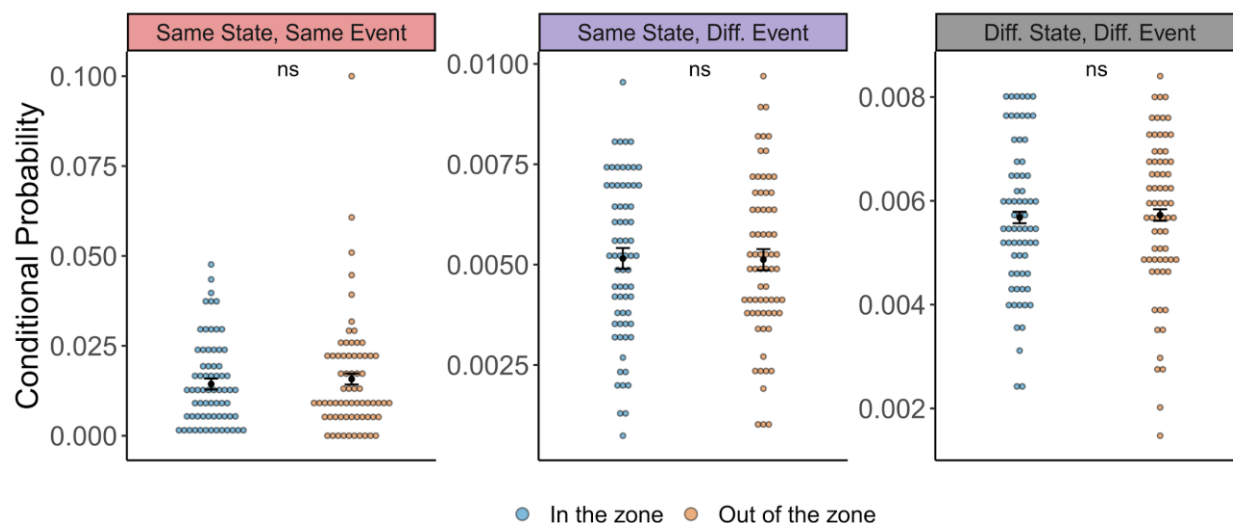
Figure 8. Recall transitions as a function of event type in Study 2. Recall transitions are shown based on whether they occurred within an "event segment' of a particular attentional state (same state, same event), across event segments of a given attentional state (same state, different event), or between attentional states (different state, different event; see Figure 2). There were no significant differences between the two attentional states in any transition type. Individual points indicate the conditional probability of each transition type (i.e., the number of times each transition type occurred divided by the number of opportunities to make a transition of that type) for each individual, separately for items encoded "in the zone" and "out of the zone". Black dots indicate the mean; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". ns = not statistically significant.

1.3.3.5 No differences in event transition types between the two attentional states

As in Study 1, we next examined recall as a function of the type of transition (**Figure 2** and **Study 1 Methods: Recall Transitions by Event Segment**). We performed a two-way repeated-measures ANOVA on the conditional probability of recall transitions with transition type (3 levels) and attentional state ("in the zone" vs. "out of the zone") as factors. We found a main effect of transition type ($F_{1.04, 69.97} = 46.76$, $p < 0.0001$, $\eta_p^2 = 0.41$, BF < 1/150). The main effect of attentional state ($F_{1,67} = 0.37$, $p = 0.54$, $\eta_p^2 = 0.006$, BF = 7.77) and the interaction between attentional state and transition type ($F_{1.05, 70.43} = 0.41$, $p = 0.53$, $\eta_p^2 = 0.006$, BF = 15.08)

were not statistically significant. Thus, each type of recall transition was not different for items encoded "in the zone" and "out of the zone" (**Figure 8**).

Given the main effect of transition type, we conducted follow-up t-tests to compare them, collapsing across the two attentional states, as we did for Study 1. We replicated Study 1 in observing that "same state, same event" transitions were significantly more likely than "same state, different event" transitions ($t_{67} = 7.03$, $p < 0.0001$, Cohen's dz $= 0.85$, 95% CI [0.007, 0.013]) and "different state, different event" transitions ($t_{67} = 6.72$, $p < 0.0001$, Cohen's dz $=$ 0.82, 95% CI [0.007, 0.012]). "Different state, different event" transitions were also significantly more likely than "same state, different event" transitions ($t_{67} = 2.34$, $p = 0.02$, Cohen's dz $= 0.28$, 95% CI [0.00008, 0.001]). This pattern is consistent with temporally organized recall, in that "same state, same event" items are the closest to one another, while "same state, different event" items are necessarily never in contiguous events (**Figure 2**). As for Study 1, we conducted a Bayesian model comparison in which we compared the two-way ANOVA reported above to a null model without an attentional state variable. This analysis revealed strong evidence in favor of the null model (BF $= 109.28$).

Thus, across these event transition analyses, we failed to find evidence for our hypotheses regarding differences between "in the zone" and "out of the zone" states.

1.3.3.6 Secondary Analysis: Faster vs. Slower RTs

As for Study 1, we examined task performance and subsequent memory with a secondary analysis in which encoding RTs were divided, by a median split, into those that were faster vs. those that were slower (see **Study 1** for procedural details). This allowed us to determine if this alternative characterization of attentional states reveals differences in recall organization that were masked by our VTC analysis, which considers RTs that are too fast or too slow to be a poor

attentional state.

There was no difference in online errors during attentional states characterized by faster vs. slower RTs ($t_{67} = 0.96$, $p = 0.34$, Cohen's dz = 0.12, 95% CI [-0.40, 1.13], BF = 4.84) and recall was also not different between these two attentional states ($t_{67} = 1.41$, $p = 0.16$, Cohen's dz = 0.18, 95% CI [-0.35, 2.07], BF = 2.92).

The lag-CRP analysis revealed only a main effect of absolute lag ($F_{11.94, 799.95} = 9.60$, $p < 0.0001$, $\eta_p^2 = 0.13$, BF < 1/150); all other main effects and interactions were not statistically significant (all ps > 0.088; all other BFs > 10.53).

The event segment analysis revealed a main effect of transition type ($F_{1.05, 70.17} = 46.80$, $p < 0.0001$, $\eta_p^2 = 0.41$, BF < 1/150), reflecting significantly more "same state, same event" transitions compared to "same state, different event" transitions ($t_{67} = 6.91$, $p < 0.0001$, Cohen's dz = 0.84, 95% CI [0.007, 0.0134]) and "different state, different event" transitions ($t_{67} = 6.88$, $p < 0.0001$, Cohen's dz = 0.83, 95% CI [0.007, 0.0127]). There was no significant difference between "different state, different event" and "same state, different event" transitions ($t_{67} = 1.83$, $p = 0.07$, Cohen's dz = 0.22, 95% CI [-0.00004, 0.001]). There was no main effect of attentional state ($F_{1,67} = 1.59$, $p = 0.21$, $\eta_p^2 = 0.023$, BF = 4.45) nor an attentional state by transition type interaction ($F_{1.03,68.96} = 1.62$, $p = 0.21$, $\eta_p^2 = 0.24$, BF = 4.90).

As for Study 1, we performed a Bayesian model comparison for the lag-CRP analysis. We compared the model reported above to a null model without an attentional state variable. This comparison revealed very strong evidence in favor of the null model (BF > 150). The same approach for the event segment analysis revealed strong evidence in favor of the null model (BF = 22.13).

Thus, like the VTC analysis, this alternative characterization of attentional states also

failed to reveal the predicted differences in the temporal organization of recall. However, unlike the VTC analysis, this approach was not successful in detecting differences in online performance based on attentional state. Thus, this alternative approach may not be as powerful in identifying fluctuating attentional states; the null effects on memory are therefore more difficult to interpret.

1.3.4 Discussion

In Study 2, we made adjustments to our task to try to encourage stronger attentional fluctuations, in particular stronger "zoning out". As before, we hypothesized that "in the zone" attentional states, vs. "out of the zone" states, aid in maintenance of temporal context representations, thus facilitating temporally organized recall. However, we did not find any evidence to support this hypothesis. While we observed differences in encoding task performance, with more errors for "out of the zone" states (replicating prior work by Esterman et al., 2013), we did not find any differences in our recall measures of interest. There was no statistically significant difference in overall recall performance for items encoded during "in the zone" vs. "out of the zone" attentional states. Furthermore, while we replicated prior work in showing both temporal contiguity and forward asymmetry effects in recall, we did not see any differences between the two attentional states in these effects. Finally, we found no differences between the two attentional states in recall transitions between different event types (**Figure 2**). Taken together, we failed to find any evidence that "in the zone" vs. "out of the zone" attentional states have a differential impact on the temporal organization of recall. This was despite our changes to study design that made it more similar to the original gradCPT procedure (see **Study 1: Discussion**).

Why did we fail to find effects of attentional states on recall organization in Study 2? One possibility is that, despite longer blocks, making semantic judgments (food vs. non-food item) may have been challenging enough to engage participants' sustained attention and hence, did not produce strong periods of "zoning out" as we hoped. In contrast, the traditional gradCPT often has participants perform a more perceptual task (e.g., male or female face). The semantic judgement may also encourage semantic clustering (Long & Kahana, 2017), which may interfere with our ability to detect differences in the temporal structure of recall between the two attentional states. Finally, it is possible that the gradual transitions between items — a core part of the gradCPT — have an unintended effect of making the task somewhat engaging, in that participants can try to identify objects at lower and lower opacities as the task goes on. Such gradual transitions also make our task less similar to standard list-learning recall tasks, which present each item in isolation. We address these limitations in Study 3.

1.4 Study 3

1.4.1 Overview

In Study 3, we made the following changes to address the limitations above. First, we changed the encoding judgement to be a perceptual one (i.e., is the image color or grayscale?) instead of a semantic one. This aligned our approach with the traditional gradCPT, which uses perceptual judgements (e.g., male or female face?). This judgement is also relatively simple and unambiguous, reducing the likelihood that RT variability may be related to task difficulty. We also made this change to ensure that the task judgement does not encourage semantic clustering of items (Long & Kahana, 2017). For example, if participants in Study 2 attempted to cluster food items together, and non-food items together, we may have had less of an opportunity to observe subtle differences in temporal clustering between the two attentional states. Second, we

removed the gradual transitions between items, and replaced this transition with a relatively long

presentation duration (3s on the screen, with a 2s inter-item interval). This change makes our

design more similar to standard memory tasks, and may additionally make the task less

engaging. Finally, in an attempt to improve recall performance, we reduced the length of our

study phase (60 items instead of the 80 in Study 2).

1.4.2 Methods

1.4.2.1 Design

1.4.2.1.1 Participants

We report data from 68 participants (M_{age} = 20.09 ± 2.20, $M_{education}$ = 14.03 ± 1.46; see

Table 1 for demographics). We do not report data from an additional 4 participants, who were

excluded due to low response rate during the encoding task (<80%, N = 1), age greater than 40

years (N = 1), recall recording issues (N = 1) and incomplete participation (N = 1). All

participants were recruited from the Columbia University participant pool and participated in an

online version of the task hosted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al.,

2020). Informed consent was obtained in accordance with the Columbia University Institutional

Review Board.

1.4.2.1.2 Stimuli

Stimuli were identical to Study 2 with the following exceptions. We chose 180 objects

from the pre-curated databases. Of these, 10% of the images (18 images) were converted to

grayscale. 90% of the images (162 images) were in color. The 180 images were divided into 3

lists of 60 images each (6 grayscale images, 54 color images). This was done manually by

ensuring there were an equal number of items from a category (e.g., tools, furniture) in each list.

1.4.2.1.3 Procedure

The procedure was identical to Study 2 with the following exceptions. In each study phase, participants viewed 60 trial-unique items from the created lists. Each image was presented for 3s followed by a fixation cross during the 2s inter-item interval (i.e., there was no gradual fading between images). Participants were asked to judge if each image was in color or grayscale: They pressed a button when it was in color, but withheld their response when it was in grayscale. Participants therefore habitually pressed one response 90% of the time. The distractor and recall phases were identical to the ones in Study 2 (**Figure 1**).

1.4.2.2 Analyses

All analyses were identical to Study 2, except that response times (RTs) were defined from the onset of the static image on each trial without any assignment algorithm (because there was no fading in).

1.4.3 Results

1.4.3.1 Defining attentional states at encoding

In the encoding task, participants viewed images and judged each as being in color (with a button press) or grayscale (by withholding their response). Overall, mean response time (RT; defined from image onset) was 1.33s (SD = 0.20). Median RT was 1.30s. As in Studies 1 and 2, we performed a variance time course analysis on the encoding phase RTs. **Figure 9A** shows the VTC analysis for one sample participant in Study 3.

The mean length of an "in the zone" segment was 3.85 trials (SD = 0.53) and the mean length of an "out of the zone" segment was 3.81 trials (SD = 0.47; Note that each trial was 5 seconds long). The mean number of fluctuations within a block (i.e., the number of times participants transitioned from one state to another) was 14.84 (SD = 2.10). The number of trials

within a segment ranged from 1 to 18 for "in the zone" states and from 1 to 20 for "out of the zone" states, across all blocks and participants.

As for prior Studies, we tested whether RT variability changed over blocks. Unlike Studies 1 and 2, however, RT variability was not significantly different across blocks ($F_{2,134} = 0.24$, $p = 0.79$, $\eta_p^2 = 0.004$). This suggests that RT variability, and the VTC approach, may not be as meaningful in this Study, perhaps because we removed the gradual image transitions that are important in the gradCPT. To further test if this is the case, we examined error rates for "in the zone" and "out of the zone" states, below.

1.4.3.2 No difference in encoding errors during "in the zone" vs "out of the zone" attentional states

Participants performed very well on the encoding task ("Is this image in color or grayscale?"). They responded to 98.80% (SD = 2.58%; Median = 100%) of the "go" color image trials, which required a response. Mean accuracy (defined as correct responses on "go" trials and withheld responses on "no-go" trials) was 97.45% (SD = 2.61%; Median = 98.34%).

We next examined errors in the encoding task as a function of attentional state. Surprisingly, we failed to replicate Studies 1 and 2: A paired samples t-test revealed that the number of encoding errors during an "out of the zone" attentional state (mean ± SD: 2.03 ± 2.40) was not significantly different from that in an "in the zone" attentional state (2.00 ± 2.34; $t_{67} = 0.10$, $p = 0.92$, Cohen's $d_z = 0.012$, 95% CI [-0.54,0.60], BF = 7.47, **Figure 9B**). Thus, unlike Studies 1 and 2, and unlike prior studies (Rosenberg et al., 2011; Esterman et al., 2013), the VTC analysis was unsuccessful in identifying fluctuations between better and worse attentional states. As in prior Studies, we next examined whether there was post-error slowing or speeding. Unlike Studies 1 and 2, we found no difference in RTs before vs. after an error (omission errors: $\beta =$

70

42.24, S.E. = 78.70, $t_{387.25}$ = 0.54, p = 0.59, 95% CI [-112.00, 196.49]; commission errors: β = -

57.49, S.E. = 83.22, $t_{138.99}$ = -0.69, p = 0.49, 95% CI [-220.61, 105.63]). This offers further

evidence that RTs in this study may not reliably index attentional fluctuations.
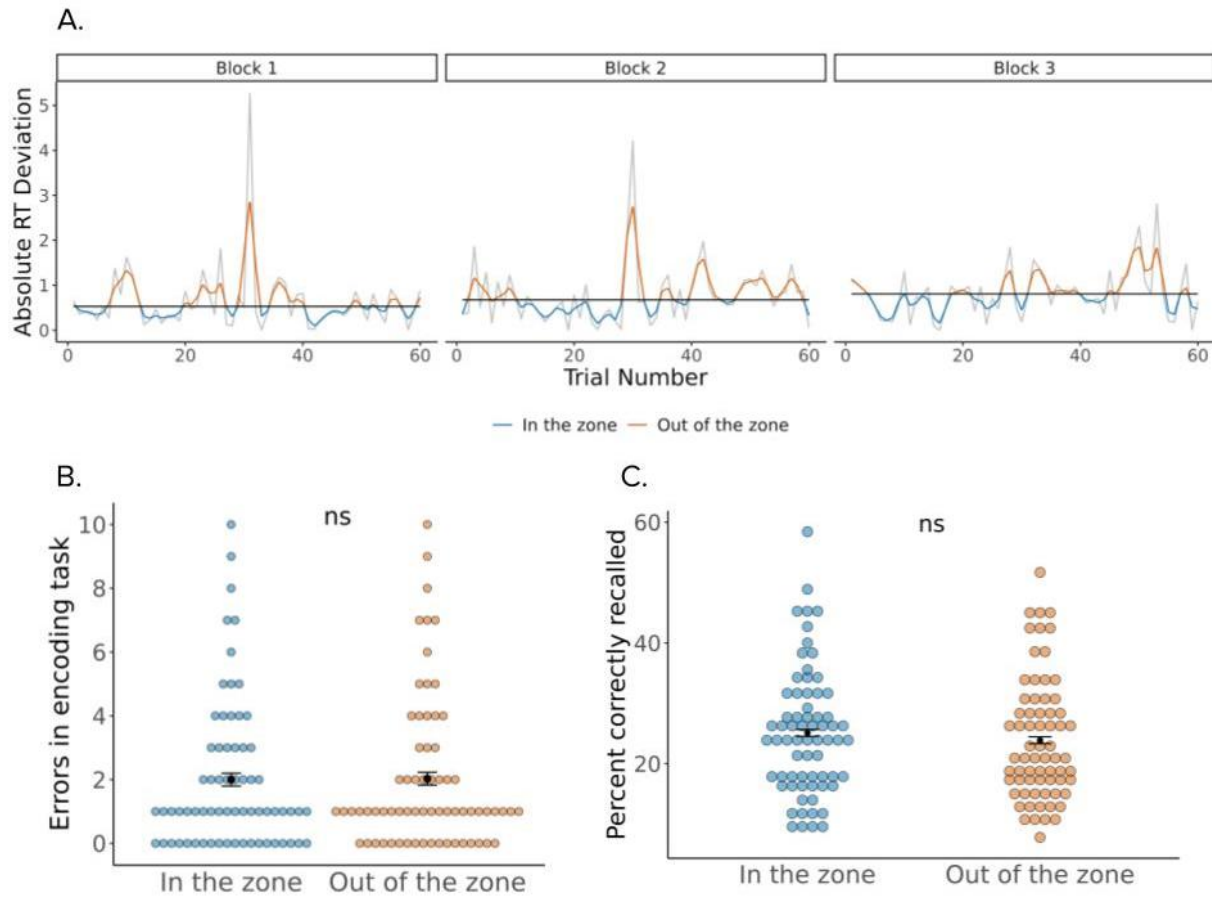
Figure 9. Encoding task performance and recall performance in Study 3. Neither encoding errors nor recall performance differ between attentional states. A. Variance Time Course (VTC) analysis for a sample participant depicting "in the zone" (blue) and "out of the zone" (orange) attentional states. Horizontal black lines indicate the median absolute RT deviation per block. Gray curves indicate raw (unsmoothed) RT deviation per block. B. Individual points show the number of encoding judgement errors made by each participant during "in the zone" and "out of the zone" attentional states. The number of encoding errors did not differ between "in the zone" vs. "out of the zone" attentional states. C. Individual points show the percentage of items correctly recalled by each participant as a function of whether items were encoded "in the zone" or "out of the zone". There was no difference in recall performance between the two states. Black points in panels B & C indicate the mean of the measure; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". ns = not statistically significant.

71

Thus, although removing gradual transitions made our design more similar to standard memory tasks, the abrupt image onsets may have captured attention, reducing "zoning out" (Esterman et al., 2013). We return to this issue in the **Discussion**. Despite this null result, we report the rest of Study 3 results for completeness.

1.4.3.3 Overall recall does not differ between the two attentional states

We next examined recall performance. Mean recall (i.e., the percentage of items correctly recalled across all blocks) was 24.32% (SD = 9.78).

We then separately examined recall performance based on whether items were encoded "in the zone" or "out of the zone". As in Studies 1 and 2, we did not find a significant difference between "in the zone" (Mean ± SD: 25.09% ± 10.40) and "out of the zone" (23.88% ± 10.10) attentional states (t_{67} = 1.51, p = 0.14, Cohen's dz = 0.18, 95% CI [-0.39, 2.81], BF = 2.56, **Figure 9C**). Thus, attentional states in this Study did not differ in either online task performance or in subsequent recall.

1.4.3.4 No differences in temporal contiguity or forward asymmetry between the two attentional states

As in Studies 1 and 2, we examined lag-CRP curves to explore the temporal organization of recall. This allowed us to determine whether the structure of memory differed between the two attentional states, even if overall memory performance did not.

As before, we constructed separate lag-CRP curves for "in the zone" vs. "out of the zone" attentional states based on successive recall of items encoded in the same state (**Figure 10B**; see **Figure 10A** and **Table 2** for overall lag-CRP). We conducted a three-way repeated-measures ANOVA with attentional state ("in the zone" vs. "out of the zone"), absolute lag (1 to 29), and direction (forward vs. backward) as factors (results summarized in **Table 3**).
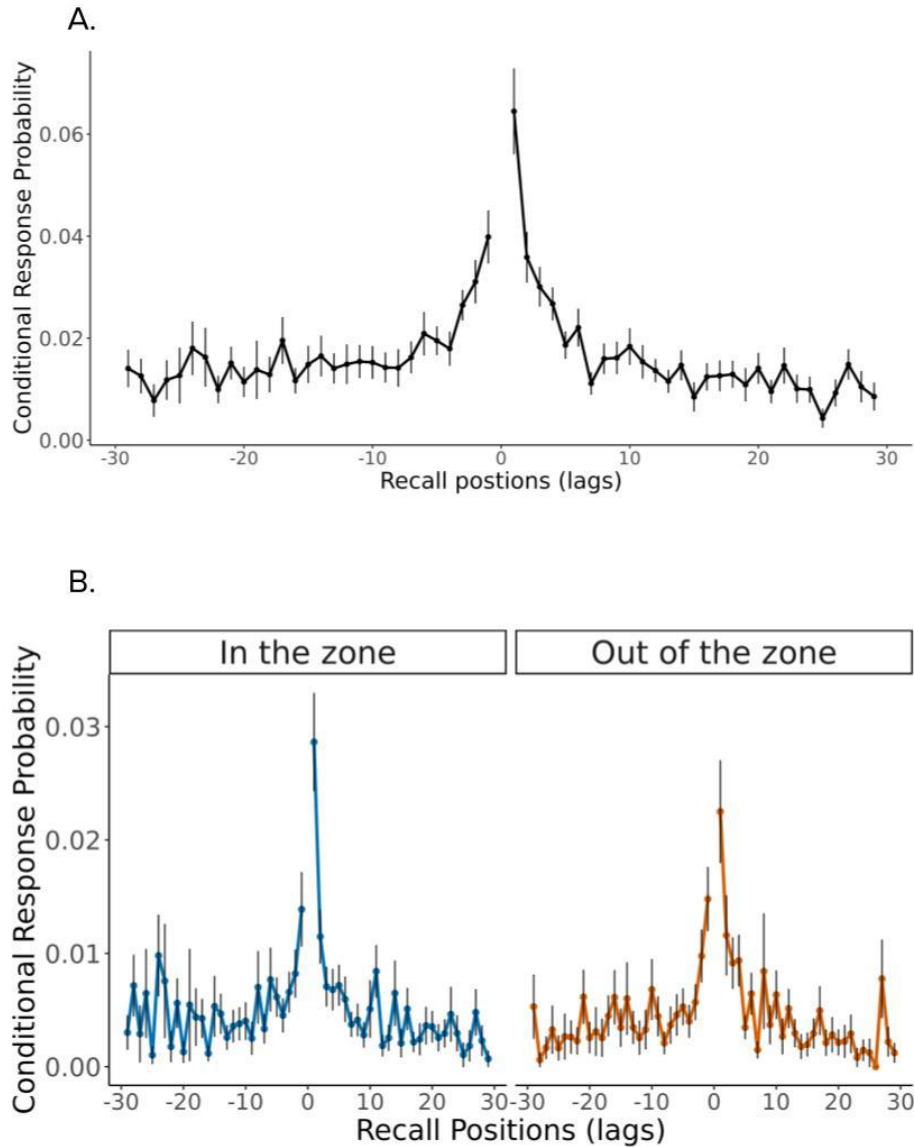
Figure 10. Lag-CRP curves overall and by attentional state for Study 3. A. Overall lag-CRP curve across participants and blocks. B. Lag-CRP curves plotted separately for items encoded "in the zone" (left) and "out of the zone" (right). There was no difference between the two attentional states in the temporal organization of recall (neither temporal contiguity nor forward asymmetry). Error bars represent the standard error.

We found a significant main effect of absolute lag ($F_{10.97,735.01}$ = 9.98, p < 0.0001, η_p^2 = 0.13, BF < 1/150): during recall, individuals were more likely to transition to items that were encoded nearby vs. farther away. There was no main effect of direction ($F_{1,67}$ = 0.066, p = 0.80,

73

$\eta_p^2 = 0.001$, BF = 38.88), but there was a significant interaction between direction and absolute lag ($F_{14.01,938.37} = 2.17$, p = 0.007, $\eta_p^2 = 0.03$, BF = 18.22): participants were more likely to recall items in the forward vs. backward direction, with this asymmetry being larger for closer vs. farther lags. There was no main effect of attentional state ($F_{1,67} = 1.68$, p = 0.20, $\eta_p^2 = 0.025$, BF = 12.39). There was also no interaction between attentional state and direction ($F_{1,67} = 0.12$, p = 0.73, $\eta_p^2 = 0.002$, BF = 26.99), no interaction between attentional state and absolute lag ($F_{14.74,987.61} = 0.82$, p = 0.66, $\eta_p^2 = 0.013$, BF > 150), nor a three-way interaction between absolute lag, direction, and attentional state ($F_{15.20,1018.34} = 0.95$, p = 0.51, $\eta_p^2 = 0.014$, BF > 150). Hence, we did not see any differences in recall organization — neither temporal contiguity nor forward asymmetry bias — based on attentional state at encoding.

As before, we conducted a follow-up analysis to examine differences between the two attentional states at the nearby lags of $\pm$ 1. From a two-way repeated-measures ANOVA with lag (+1 vs. -1) and attentional state ("in the zone" vs. "out of the zone") as factors, we found only a significant main effect of lag ($F_{1,67} = 12.57$, p = 0.0007, $\eta_p2 = 0.16$, BF = 0.028). The main effect of attentional state ($F_{1,67} = 0.55$, p = 0.46, $\eta_p2 = 0.008$, BF = 6.36), and the interaction between attentional state and lag, was not statistically significant ($F_{1,67} = 1.32$, p = 0.26, $\eta_p2 = 0.019$, BF = 3.08). The significant main effect of lag (+1 vs. -1) confirms that individuals are more likely to make forward vs. backward transitions at the closest lag. However, forward asymmetry at the $\pm$ 1 lags was not different between the two states.

As for prior Studies, we conducted a Bayesian model comparison in which we compared the three-way ANOVA reported above to a null model without an attentional state variable. Once again, this analysis revealed very strong evidence in favor of the null model (BF > 150).

We therefore once again replicated the finding that recall is temporally organized. However, the temporal organization of recall was not different between the two attentional states. This replicates the null findings from the lag-CRP analyses in Studies 1 and 2.

1.4.3.5 No differences in event transition types between the two attentional states

As in Studies 1 and 2, we next examined recall transitions as a function of the type of event segment (see Figure 2 and Study 1 Methods: Recall Transitions by Event Segment). To do this, we performed a two-way repeated-measures ANOVA with transition type (3 levels) and attentional state ("in the zone" vs. "out of the zone") as factors. As before, we hypothesized that "same state, same event" and "same state, different event" transitions may be more likely for items encoded "in the zone" vs. "out of the zone".

We found a main effect of transition type ($F_{1.06,70.93} = 34.19$, $p < 0.0001$, $\eta_p2 = 0.34$, BF $<$ 1/150). The main effect of attentional state ($F_{1,67} = 2.30$, $p = 0.13$, $\eta_p2 = 0.03$, BF $= 4.41$), and the interaction between attentional state and transition type ($F_{1.14,76.32} = 1.32$, $p = 0.26$, $\eta_p2 = 0.02$, BF $= 10.66$), were not statistically significant. Thus, each type of recall transition was not differentially likely for items encoded "in the zone" and "out of the zone" (Figure 11).

As for prior Studies, we conducted follow-up t-tests to understand the main effect of transition types, collapsing across the two attentional states. We found that: 1) "same state, same event" transitions were significantly more likely compared to "same state, different event" transitions ($t_{67} = 6.10$, $p < 0.0001$, Cohen's dz $= 0.74$, 95% CI [0.008, 0.015]) and "different state, different event" transitions ($t_{67} = 5.65$, $p < 0.0001$, Cohen's dz $= 0.69$, 95% CI [0.006, 0.013]), and 2) "different state, different event" transitions were significantly more likely compared to "same state, different event" transitions ($t_{67} = 3.47$, $p = 0.0009$, Cohen's dz $= 0.42$, 95% CI [0.0005, 0.002]).
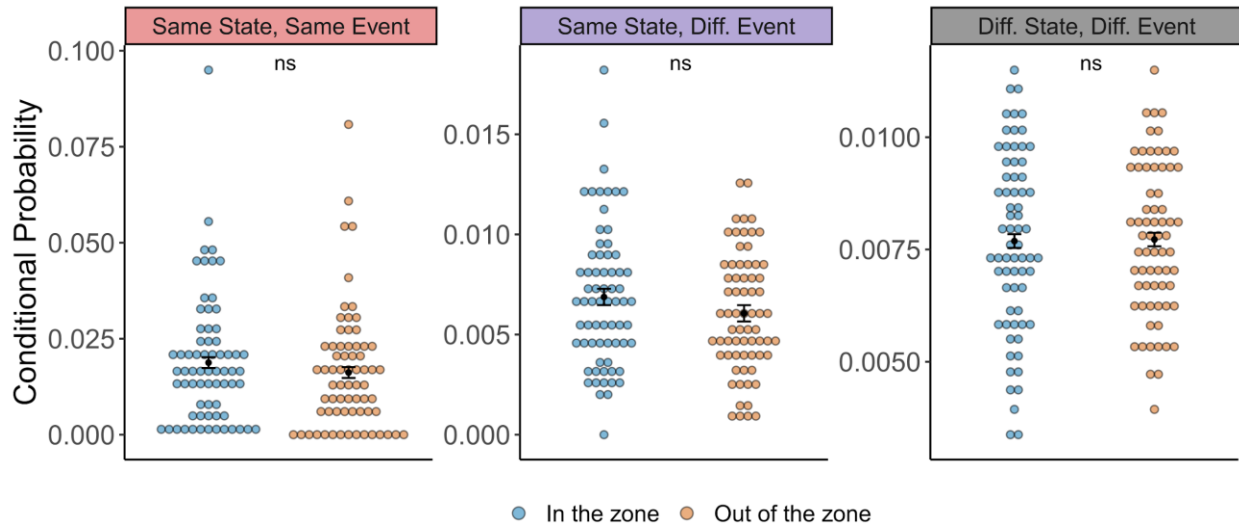
Figure 11. Recall transitions as a function of event type in Study 3. Recall transitions are shown based on whether they occurred within an "event segment' of a particular attentional state (same state, same event), across event segments of a given attentional state (same state, different event), or between attentional states (different state, different event; see Figure 2). There were no significant differences in any transition type between the two attentional states. Individual points indicate the conditional probability of each transition type (i.e., the number of times each transition type occurred divided by the number of opportunities to make a transition of that type) for each individual, separately for items encoded "in the zone" and "out of the zone". Black dots indicate the mean; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". ns = not statistically significant.

As for prior Studies we conducted a Bayesian model comparison in which we compared the two-way ANOVA reported above to a null model without an attentional state variable. This analysis again revealed strong evidence in favor of the null model (BF = 44.01). As in prior Studies, therefore, we did not find any evidence that recall transition types differed between the two attentional states.

1.4.3.6 Secondary Analysis: Faster vs. Slower RTs

As before, we examined task performance and subsequent memory with a secondary analysis. Rather than using the VTC analysis to characterize "in the zone" and "out of the zone" states, we divided encoding RTs, by a median split, into those that were faster vs. those that were

slower. We found no difference in error rates during attentional states characterized by faster vs. slower RTs (t_{67} = 0.85, p = 0.40, Cohen's dz = 0.10, 95% CI [-0.32, 0.79], BF = 5.30). This is consistent with Studies 1 and 2, in which this analysis also failed to reliably predict online errors, and also consistent with the VTC analysis above, which likewise failed to find differences in error rates for "in the zone" vs. "out of the zone" states in this Study.

This analysis did, however, reveal a difference in subsequent recall performance: attentional states associated with slower (vs. faster) RTs at encoding were associated with better subsequent recall (t_{67} = 2.18, p = 0.033, Cohen's dz = 0.26, 95% CI [0.14, 3.13], BF = 0.82). The lag-CRP analysis nevertheless failed to reveal differences in the temporal organization of recall: there was only a main effect of absolute lag ($F_{11.90, 797.25}$ = 9.23, p < 0.0001, η_p^2 = 0.12, BF < 1/150; all other ps > 0.06; all other BFs > 24.81).

The event segment analysis revealed a main effect of transition type ($F_{1.04, 69.80}$ = 29.90, p < 0.0001, η_p^2 = 0.31, BF < 1/150), reflecting significantly more "same state, same event" transitions compared to "same state, different event" transitions (t_{67} = 5.55, p < 0.0001, Cohen's dz = 0.67, 95% CI [0.007, 0.015]) and "different state, different event" transitions (t_{67} = 5.46, p < 0.0001, Cohen's dz = 0.66, 95% CI [0.007, 0.015]). There was no significant difference between "different state, different event" and "same state, different event" transitions (t_{67} = 1.50, p = 0.14, Cohen's dz = 0.18, 95% CI [-0.0002, 0.001]). There was no main effect of attentional state ($F_{1,67}$ = 2.55, p = 0.11, η_p^2 = 0.04, BF = 3.60) nor an attentional state by transition type interaction ($F_{1.12,74.90}$ = 1.55, p = 0.22, η_p^2 = 0.02, BF = 8.60).

As for Studies 1 and 2, we performed a Bayesian model comparison for the lag-CRP analysis, in which we compared the model reported above to a null model without an attentional state variable. This comparison revealed very strong evidence in favor of the null model (BF >

150). The same approach for the event segment analysis revealed strong evidence in favor of the null model (BF = 31.77).

Thus, like the VTC analysis, this alternative characterization of attentional states also failed to reveal statistically robust differences in the temporal organization of recall.

1.4.4 Discussion

In Study 3, we made further adjustments to our task to encourage more "zoning out" and to make our procedure more similar to standard list-learning recall tasks. We again tested the hypothesis that "in the zone", vs. "out of the zone", states aid in maintenance of temporal context representations, thus encouraging temporally organized recall. However, we did not find any evidence to support this hypothesis.

Contrary to Studies 1 and 2, we did not observe any differences in encoding task performance across the two attentional states. Why might this be the case? First, we removed the gradual transitions between images. Although this change made our task more similar to standard list-learning recall tasks, the abrupt image onsets could have captured attention, thus preventing participants from "zoning out" (Esterman et al., 2013; Rosenberg et al., 2013). Second, long and variable inter-stimulus durations result in more attentional lapses than short, fixed durations (Unsworth et al., 2018). Our short and fixed ITI (2 seconds) may therefore have made it less likely that participants would zone out. Finally, we changed the encoding judgement to a simple perceptual judgement. Although this change made our task more similar to the traditional gradCPT, which often uses perceptual judgments, it had the effect of increasing the accuracy of encoding judgements. Having such a low error rate (~2 errors, on average, per participant in each attentional state) may have hurt our chances of seeing differences between the two states. This low error rate may have been due to the saliency of color changes, which might capture attention

even if a participant had previously been zoning out. The lack of a difference in errors between

"in the zone" and "out of the zone" states makes it difficult to interpret the lack of differences

between them in subsequent recall organization: we may not have adequately captured

attentional fluctuations based on RT. This prompted us to conduct Study 4, which is similar to

Study 2, with some minor changes to once again try to induce stronger attentional fluctuations.

1.5 Study 4

1.5.1 Overview

In Study 3, online task performance was not different between "in the zone" and "out of

the zone" states. This makes the null effects in recall organization difficult to interpret, because

we may not have successfully differentiated between better and worse attentional states. We

therefore conducted another study with the gradual transitions used in Studies 1 and 2, in which

we successfully replicated prior work showing more errors during "out of the zone" states

(Esterman et al., 2013; Rosenberg et al., 2013). Study 4 used a similar design as Study 2 (go/no-

go procedure, with responses for non-food items and withheld responses for food items). Two

changes were made: first, there were 2 blocks of 120 images each (instead of 3 blocks of 80

images each), and second, the trial duration (from the onset of an image fading in until it became

100% clear) was reduced from 6s to 4s. Both of these changes were implemented to bring our

design closer to the traditional gradCPT, which typically uses fast presentation durations and

many trials. In this way, we hoped to induce stronger attentional fluctuations. The increase in

block length, however, means that our procedure deviates even more from standard list-learning

recall studies, which typically use short lists (Murdock Jr., 1962; Cortis et al., 2015; also see

Healey et al., 2019). We return to this point in the General Discussion.

1.5.2 Methods

1.5.2.1 Design

1.5.2.1.1 Participants

Pilot data using the Study 4 procedure revealed that participants had worse recall performance than our earlier studies (likely because Study 4 blocks were longer than those in our prior studies). We therefore opted to collect a larger sample size, so that summed recall performance across all participants would be comparable to Study 2 (Baker et al., 2020). We report data from 124 participants (M_{age} = 21.42 ± 6.26, $M_{education}$ = 13.63 ± 1.54; see Table 1 for demographics). We do not report data from an additional 32 participants, who were excluded due to image loading errors (N = 1), low response rate during the encoding task (<80%, N = 20), recall recording issues (N = 10), and no recall (N = 1). Of the final sample, 5 participants were recruited through Prolific (www.prolific.co) and the rest (119 participants) were recruited from the Columbia University participant pool. All participants completed an online version of the task hosted on the Gorilla platform (www.gorilla.sc; Anwyl-Irvine et al., 2020). Informed consent was obtained in accordance with the Columbia University Institutional Review Board.

1.5.2.1.2 Stimuli

Stimuli were identical to Study 2, except that the 240 images were divided into 2 lists of 120 images each (12 food, 108 non-food images).

1.5.2.1.3 Procedure

The procedure was identical to Study 2 with the following exceptions. The experiment consisted of 2 blocks, each of which included a study phase, a distractor phase, and a recall phase (Figure 1). In each study phase, participants viewed 120 trial-unique items, which transitioned slowly from one into another. Trial duration was 4s instead of 6s.

The distractor phase was identical to that in Study 2. The recall phase was similar to Study 2, except that participants were given the option to recall for a longer duration. Participants were initially given 4 minutes (broken into 2 recordings of 2 minutes each) to verbally recall items from the study phase. After the initial 4 minutes of recording, participants were given the option of recording for an additional 2 minutes. This was done because the encoding blocks in Study 4 were longer than those in Study 2; thus, we wanted to give participants more time to recall if they needed it.

1.5.2.2 Analyses

Analyses were identical to Study 2.

1.5.3 Results

1.5.3.1 Defining attentional states at encoding

In the encoding task, participants viewed images and judged each as being a non-food item (with a button press) or a food item (by withholding their response). Overall, mean response time (RT; defined from image onset) was 2.53s (SD = 0.36). Median RT was 2.55s.

As in Studies 1-3, we performed a variance time course analysis on the encoding phase RTs. **Figure 12A** shows the VTC analysis for one sample participant in Study 4. The mean length of an "in the zone" segment was 4.97 trials (SD = 0.58) and the mean length of an "out of the zone" segment was 4.93 trials (SD = 0.56; Note that each trial was 4 seconds long). The mean number of fluctuations within a block (i.e., the number of times participants transitioned from one state to another) was 23.45 (SD = 2.67). The number of trials within a segment ranged from 1 to 33 for "in the zone" states and from 1 to 27 for "out of the zone" states, across all blocks and participants.
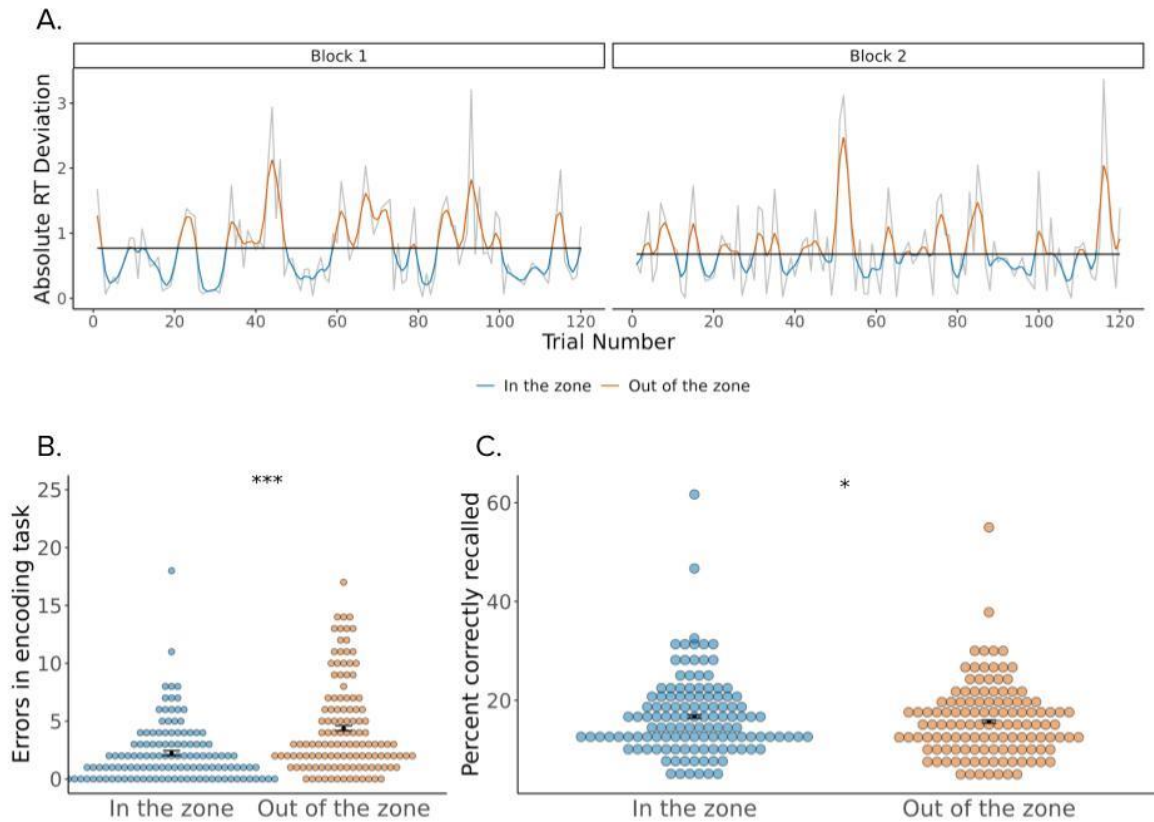
Figure 12. Encoding task performance and recall performance in Study 4. Both encoding errors and recall performance differ between the two attentional states. A. Variance Time Course (VTC) analysis for a sample participant, depicting "in the zone" (blue) and "out of the zone" (orange) attentional states. Horizontal black lines indicate the median absolute RT deviation per block. Gray curves indicate raw (unsmoothed) RT deviation per block. B. Individual points show the number of encoding judgment errors made by each participant during "in the zone" and "out of the zone" attentional states. Participants made significantly more encoding errors during the "out of the zone" state. C. Individual points show the percentage of items correctly recalled by each participant as a function of whether items were encoded "in the zone" or "out of the zone". There was a small but statistically significant difference in recall performance: recall was higher for items encoded while "in the zone" vs. "out of the zone". Black points in panels B & C indicate the mean of the measure; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". * $p < 0.05$, *** $p < .0001$.

As for prior Studies, we performed an initial validation check for the VTC approach by testing whether RT variability changed over blocks. As in Studies 1 and 2, RT variability increased across blocks ($t_{123} = 6.79$, $p < 0.0001$, Cohen's $dz = 0.61$, 95% CI [68.22, 124.34]),

which may be suggestive of increased fatigue or loss of focus as the experiment continued. This supports the use of RT variability to differentiate "in the zone" and "out of the zone" states. These attentional states were used to examine accuracy on the encoding task and subsequent recall performance, described below.

1.5.3.2 More encoding errors during "out of the zone" attentional states

Participants once again performed very well on the encoding task ("Is this item a food or a non-food item?"). They responded to 99.16% (SD = 2.10%; Median = 99.54%) of the "go" non-food image trials, which required a response. Mean accuracy (defined as correct responses on "go" trials and withheld responses on "no-go" trials) was 97.08% (SD = 2.55%; Median = 97.92%).

We next examined errors in the encoding task as a function of attentional state. We replicated Studies 1 and 2 and prior studies (e.g., Esterman et al., 2013, 2014; Rosenberg et al., 2013, 2015; Fortenbaugh et al., 2018). Participants made significantly more errors during an "out of the zone" attentional state (mean ± SD: 4.39 ± 4.00) compared to an "in the zone" attentional state (2.23 ± 2.63; t_{123} = 7.04, p < 0.001, Cohen's dz = 0.63, 95% CI [1.55, 2.76], BF < 1/150, **Figure 12B**). Thus, the VTC analysis was once again successful in identifying fluctuations between better and worse attentional states.

As in prior Studies, we performed an additional validation check of whether RTs are sensitive to attentional fluctuations in our task, by comparing RTs before vs. after an error. As in Study 2, participants were faster to respond after vs. before an omission error (β = -970.21, S.E. = 67.82, $t_{650.25}$ = -14.31, p < 0.0001, 95% CI [-1103.13, -837.29]). Conversely, participants were slower to respond after vs. before a commission error (β = 210.90, S.E. = 62.80, $t_{461.06}$ = 3.36, p = 0.0009, 95% CI [87.82, 333.98]). These results are consistent with participants speeding up after

realizing they failed to respond to the previous trial, and slowing down after realizing they

incorrectly responded when they should not have. Thus, post-error slowing after commission

errors and post-error speeding after omission errors show that RTs can index attentional

fluctuations in our task.

1.5.3.3 Recall performance is better for items encoded during "in the zone" states

We next examined recall performance. Mean recall (i.e., the percentage of items correctly

recalled across all blocks) was 16.26% (SD = 7.36).

We then separately examined recall performance based on whether items were encoded

"in the zone" or "out of the zone". Unlike Studies 1-3, we found a significant difference between

the attentional states, such that participants recalled a higher percentage of items that were

encoded during an "in the zone" (Mean ± SD: 16.72% ± 8.12) vs. "out of the zone" (15.69% ±

7.51) attentional state (t_{123} = 2.21, p = 0.03, Cohen's dz = 0.20, 95% CI [0.11, 1.95], BF = 0.96,

Figure 12C). Thus, attentional states at encoding were associated with a difference in both

online task performance and subsequent recall performance, with both being superior for "in the

zone" states. This suggests that the VTC analysis was successful in identifying better vs worse

attentional states.

1.5.3.4 No differences in temporal contiguity or forward asymmetry between the
two attentional states

As in prior Studies, we examined lag-CRP curves to explore the temporal organization of

recall. This allowed us to determine whether the structure of memory differed between the two

attentional states.

As before, we constructed separate lag-CRP curves for "in the zone" vs. "out of the zone"

attentional states based on successive recall of items encoded in the same state (**Figure 13B**; see

Figure 13A and **Table 2** for overall lag-CRP). We conducted a three-way repeated-measures ANOVA with attentional state ("in the zone" vs. "out of the zone"), absolute lag (1 to 29), and direction (forward vs. backward) as factors (results summarized in **Table 3**).

We found a significant main effect of absolute lag ($F_{10.22, 1256.64} = 10.33$, $p < 0.0001$, $\eta_p^2 = 0.08$, BF < 1/150): during recall, individuals were more likely to transition to items that were encoded nearby vs. farther away. There was no main effect of direction ($F_{1,123} = 1.69$, $p = 0.20$, $\eta_p^2 = 0.014$, BF = 24.95). However, there was a significant interaction between direction and absolute lag ($F_{12.59, 1549.09} = 2.00$, $p = 0.02$, $\eta_p^2 = 0.016$, BF > 150). This interaction arose because forward asymmetry (forward > backward recall) was more pronounced for closer vs. farther lags. Unlike Studies 1-3, but consistent with the recall difference between attentional states, there was a main effect of attentional state ("in the zone" > "out of the zone"; $F_{1,123} = 3.97$, $p = 0.0485$, $\eta_p^2 = 0.03$, BF = 2.08). However, as in Studies 1-3, there was no interaction between attentional state and direction ($F_{1,123} = 0.52$, $p = 0.47$, $\eta_p^2 = 0.004$, BF = 29.46), no interaction between attentional state and absolute lag ($F_{1313, 1615.28} = 0.62$, $p = 0.84$, $\eta_p^2 = 0.005$, BF > 150), nor a three-way interaction between absolute lag, direction, and attentional state ($F_{13.16, 1618.76} = 0.55$, $p = 0.90$, $\eta_p^2 = 0.0045$, BF > 150). Hence, we did not see any differences in recall organization — neither temporal contiguity nor forward asymmetry bias — based on attentional state at encoding.

As before, we also conducted a follow-up analysis to examine differences between the two attentional states at the nearby lags of ± 1. From a two-way repeated-measures ANOVA with lag (+1 vs. -1) and attentional state ("in the zone" vs. "out of the zone") as factors, we found only a significant main effect of lag ($F_{1,123} = 9.69$, $p = 0.002$, $\eta_p^2 = 0.07$, BF = 0.043). The main effect of attentional state ($F_{1,123} = 0.30$, $p = 0.59$, $\eta_p^2 = 0.002$, BF = 8.96), and the interaction between attentional state and lag, was not statistically significant ($F_{1,123} = 0.01$, $p = 0.92$, $\eta_p^2 =$

0.00008, BF = 7.20). The significant main effect of lag (+1 vs. -1) confirms that individuals are more likely to make forward vs. backward transitions at the closest lag. Nevertheless, forward asymmetry at the ± 1 lags was not different between the two states.
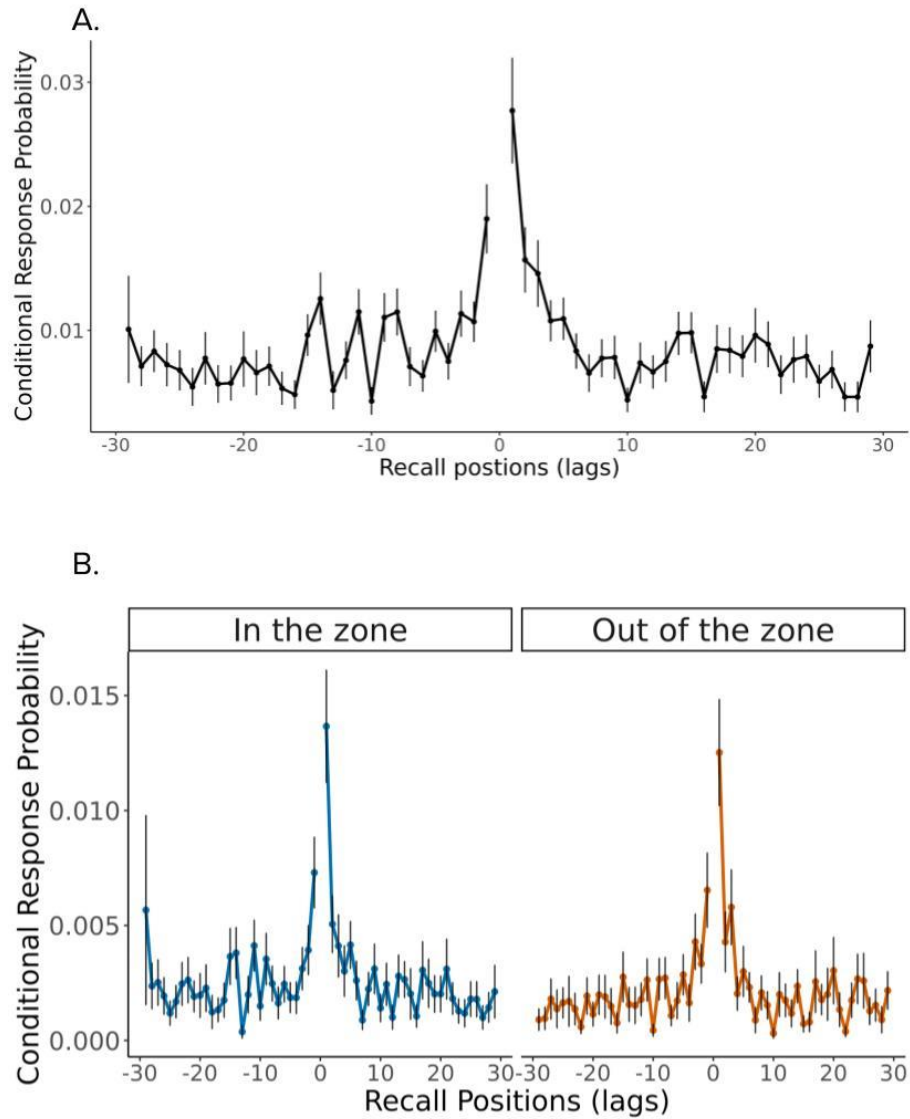
Figure 13. Lag-CRP curves overall and by attentional state for Study 4. A. Overall lag-CRP curve across participants and blocks. B. Lag-CRP curves plotted separately for items encoded "in the zone" (left) and "out of the zone" (right). There was no difference between the two attentional states in the temporal organization of recall (neither temporal contiguity nor forward asymmetry). Error bars represent the standard error.

As for prior Studies, we conducted a Bayesian model comparison in which we compared the three-way ANOVA reported above to a null model without an attentional state variable. Once again, this analysis revealed very strong evidence in favor of the null model (BF > 150). We therefore once again replicated the finding that recall is temporally organized. However, the temporal organization of recall was not different between the two attentional states. This replicates the null findings from the lag-CRP analyses in Studies 1-3. This null effect was observed even though overall recall was higher for items encoded "in the zone" vs "out of the zone".

1.5.3.5 No differences in event transition types between the two attentional states

We next examined recall transitions as a function of the type of event segment (see **Figure 2** and **Study 1 Methods: Recall Transitions by Event Segment**).

To do this, we performed a two-way repeated-measures ANOVA with transition type (3 levels) and attentional state ("in the zone" vs. "out of the zone") as factors. As before, we hypothesized that "same state, same event" and "same state, different event" transitions may be more likely for items encoded "in the zone" vs. "out of the zone".

We found a main effect of transition type ($F_{1.09, 134.05} = 30.45$, $p < 0.0001$, $\eta_p^2 = 0.20$, BF < 1/150). The main effect of attentional state ($F_{1,123} = 1.23$, $p = 0.27$, $\eta_p^2 = 0.01$, BF = 6.93), and the interaction between attentional state and transition type ($F_{1.14, 140.14} = 0.36$, $p = 0.58$, $\eta_p^2 = 0.003$, BF = 25.54), were not statistically significant. This suggests that each type of recall transition is not differentially likely for items encoded "in the zone" and "out of the zone" (**Figure 14**).
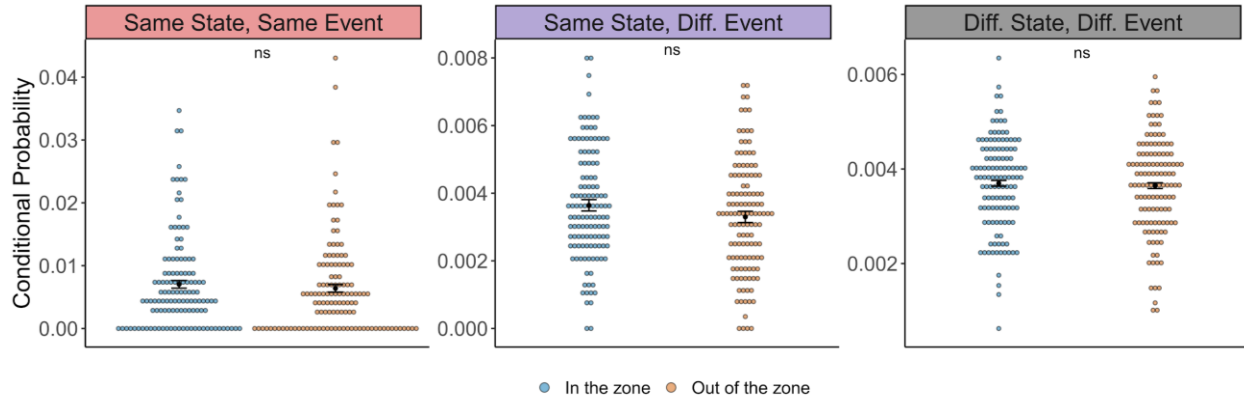
Figure 14. Recall transitions as a function of event type in Study 4. Recall transitions are shown based on whether they occurred within an "event segment' of a particular attentional state (same state, same event), across event segments of a given attentional state (same state, different event), or between attentional states (different state, different event; see Figure 2). There were no significant differences in any transition type between the two attentional states. Individual points indicate the conditional probability of each transition type (i.e., the number of times each transition type occurred divided by the number of opportunities to make a transition of that type) for each individual, separately for items encoded "in the zone" and "out of the zone". Black dots indicate the mean; error bars indicate the standard error of the within-participant difference between "in the zone" and "out of the zone". ns = not statistically significant.

We conducted follow-up t-tests to understand the main effect of transition types, collapsing across the two attentional states. As before, we found that "same state, same event" transitions were significantly more likely compared to "same state, different event" transitions (t_{123} = 5.75, p < 0.0001, Cohen's dz = 0.52, 95% CI [0.002, 0.004]) and "different state, different event" transitions (t_{123} = 5.43, p < 0.0001, Cohen's dz = 0.49, 95% CI [0.002, 0.004]), There was no significant difference between "different state, different event" transitions and "same state, different event" transitions (t_{123} = 1.46, p = 0.15, Cohen's dz = 0.13, 95% CI [-0.00007, 0.0005]).

As for prior Studies we conducted a Bayesian model comparison in which we compared the two-way ANOVA reported above to a null model without an attentional state variable. This analysis once again revealed very strong evidence in favor of the null model (BF > 150). Thus,

as in prior Studies, we found no evidence that recall transitions differed between the two attentional states.

1.5.3.6 Secondary Analysis: Faster vs. Slower RTs

As before, we examined task performance and subsequent memory with a secondary analysis in which encoding RTs were divided, by a median split, into those that were faster vs. those that were slower. Participants made more errors during attentional states characterized by faster vs. slower RTs ($t_{123} = 2.04$, $p = 0.043$, Cohen's $dz = 0.18$, 95% CI [0.02, 1.06], BF = 1.35). They also exhibited worse subsequent recall for items associated with faster vs. slower RTs at encoding ($t_{123} = 2.18$, $p = 0.03$, Cohen's $dz = 0.20$, 95% CI [0.09, 1.87], BF = 1.02). These results are conceptually similar to those with the VTC analysis, in which more errors and worse recall were linked to "out of the zone" vs. "in the zone" states.

The lag-CRP analysis, however, failed to reveal any differences in the temporal organization of recall: there was only a main effect of absolute lag ($F_{15.56,\ 1914.03} = 6.46$, $p < 0.0001$, $\eta_p^2 = 0.049$, BF < 1/150; all other ps > 0.24; all other BFs > 26.92).

The event segment analysis revealed a main effect of transition type ($F_{1.03,126.94} = 32.68$, $p < 0.0001$, $\eta_p^2 = 0.21$, BF < 1/150), reflecting significantly more "same state, same event" transitions compared to "same state, different event" transitions ($t_{123} = 5.72$, $p < 0.0001$, Cohen's $dz = 0.51$, 95% CI [0.003, 0.007]) and "different state, different event" transitions ($t_{123} = 5.78$, $p < 0.0001$, Cohen's $dz = 0.52$, 95% CI [0.003, 0.007]). There was no significant difference between "different state, different event" and "same state, different event" transitions ($t_{123} = 0.74$, $p = 0.46$, Cohen's $dz = 0.07$, 95% CI [-0.0002, 0.0004]). There was no main effect of attentional state ($F_{1,123} = 0.085$, $p = 0.77$, $\eta_p^2 = 0.0007$, BF = 11.26) nor an attentional state by transition type interaction ($F_{1.05,129.40} = 0.065$, $p = 0.81$, $\eta_p^2 = 0.0005$, BF = 32.97).

Finally, as for prior Studies, we performed a Bayesian model comparison between the lag-CRP model reported above and a null model without an attentional state variable. This comparison revealed very strong evidence in favor of the null model (BF > 150). Likewise, the same approach for the event segment analysis revealed very strong evidence in favor of the null model (BF > 150).

Thus, as for other Studies, there was no statistically significant difference in recall organization between attentional states associated with faster vs. slower RTs at encoding. Our main conclusions — lack of evidence that attentional states at encoding influence the temporal organization of memory — therefore hold with this alternative characterization of attentional states.

1.5.4 Discussion

In Study 4, we sought to replicate the findings from our prior Studies, particularly the lack of a difference in the temporal structure of recall for items encoded "in the zone" vs. "out of the zone". We used a design similar to Study 2 but increased block length and reduced stimulus presentation time to make our design more similar to the traditional gradCPT. As in our other Studies, we failed to find any evidence for more temporally structured recall for "in the zone" vs "out of the zone" encoding states.

Interestingly though, this was the first Study in which we found an effect of "in the zone" vs. "out of the zone" attentional states on subsequent recall: participants recalled more items encoded during an "in the zone" state compared to an "out of the zone" state. Thus, although we found fewer online errors and better overall recall for items encoded "in the zone" vs. "out of the zone" (Figure 12), and although we replicated temporal contiguity effects in overall recall, we still failed to find evidence for differences across attentional states in recall organization. This

suggests that our VTC analysis was able to successfully differentiate between better and worse attentional states, but these states were remarkably similar in the temporal organization of recall. We discuss the implications of our findings and their relation to prior work in the General Discussion.

1.6 General Discussion

1.6.1 Summary of findings

We examined the behavioral effects of endogenous fluctuations in attention on the temporal organization of memory. We used response time variability at encoding to characterize two attentional states: the relatively good "in the zone" state and the relatively worse "out of the zone" state. We hypothesized that good (vs. bad) attentional states at encoding will be more conducive to maintaining temporal context representations, thus promoting more temporally organized recall and facilitating "leaps" between temporally distant but cognitively similar attentional states. However, across four Studies we failed to find evidence to support either hypothesis. Indeed, Bayesian comparisons — between models that included attentional state as a variable and null models that did not — consistently revealed strong evidence for the null hypothesis that attentional states at encoding do not influence the temporal organization of recall.

We replicated previous findings that individuals make more errors in online task performance during "out of the zone" states (e.g., Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014; Rosenberg et al., 2015; Fortenbaugh et al., 2018). In Study 4, we also found that recall was worse when encoding occurred in an "out of the zone" state. We also replicated several well-established memory phenomena, including temporal contiguity effects and forward asymmetry in recall (Kahana, 1996; Howard & Kahana, 2002a; Healey et al., 2019). Despite this, we found no evidence that the temporal organization of recall was affected by

91

attentional fluctuations at encoding: recall was robustly temporally organized, even when encoding occurred in relatively poor attentional states. Furthermore, even when we conducted an analysis that combined Studies 1, 2, and 4 (our diagnostic experiments, in which "out of the zone" attentional states were associated with more online errors), we found no evidence of differential temporal organization of recall as a function of encoding attentional state (all ps > 0.29 for interactions involving attentional state). Yet, there was strong evidence for temporally organized recall generally (main effect of lag: $F_{11.21,2871.99} = 24.78$, p < 0.0001, $\eta_p^2 = 0.09$; lag by direction interaction: $F_{11.40,2918.07} = 2.49$, p = 0.0036, $\eta_p^2 = 0.01$). Together, our findings suggest that temporal context serves as a strong scaffold for episodic memory, one that can overcome spontaneous fluctuations in attentional states. Furthermore, our Studies highlight the difficulty of merging tasks of sustained attention and memory recall, and the numerous factors that must be considered when doing so (e.g., list length, trial-unique items, repetitive responses). We explore other potential reasons for our findings, their implications, and recommendations for future research in the sections below.

1.6.2 Exploring reasons for the null effect of attentional states on recall organization

Why did we not see the hypothesized relationship between attentional states and temporal organization of recall? One possibility is that converting our experiments to online studies increased the noise in our data, hence obscuring any potential effects. However, data from Study 1 suggests this isn't the case: in control analyses, we found no differences in any measures of interest between online and in-person participants. Furthermore, we replicated established in-lab phenomena in our online-only studies (Studies 2-4) such as the temporal contiguity effect and forward asymmetry in free recall, as well as more errors for "out of the zone" attentional states.

Thus, it is unlikely that moving to online experiments was the main reason behind the lack of evidence supporting our hypothesis.

A second possibility is that measures of RT variability are not sensitive to spontaneous fluctuations in attentional states, and thus, we failed to characterize these states. However, there is strong evidence from sustained attention studies that response time variability effectively captures subtle fluctuations in attentional states, which can then be related to online task performance (e.g., Robertson et al., 1997; Esterman et al., 2013, 2014; Rosenberg et al., 2013). Furthermore, other studies have used RT variability-based attentional states as a trait-level measure and related it to episodic memory (Madore et al., 2020). Our results were consistent with these effects: compared to "in the zone" states, "out of the zone" attentional states were associated with more errors during the encoding task and, in Study 4, worse recall. Furthermore, we found that RTs reliably fluctuated around errors, and that RT variability tended to increase over the course of the experiment. These findings together suggest that RTs, and RT variability, capture important cognitive phenomena in our tasks.

Despite the success of the variance time course (VTC) analysis that we focused on in the current study (Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014; Rosenberg et al., 2015; Madore et al., 2020), there are alternative ways to quantify better vs worse attentional states. For example, some studies of sustained attention have shown that faster (vs. slower) RTs are linked to more online errors and worse subsequent recognition memory (e.g., Robertson et al., 1997; Cheyne et al., 2006; McVay & Kane, 2012; deBettencourt et al., 2018; Wakeland-Hart et al., 2022). Conversely, other studies link longer eye fixations and slower response times to mind wandering and attentional lapses (e.g., Smallwood et al., 2003; Cheyne et al., 2009; Feng et al., 2013; Yanko & Spalek, 2013; Kam & Handy, 2014; Henríquez et al., 2016; Krasich et al.,

2018; Zhang et al., 2021). Thus, some studies suggest that slower (vs. faster) RTs are associated with better attentional states; other studies suggest the opposite. Both of these approaches can be contrasted to the VTC analysis, in which RTs that are too fast or too slow (relative to the mean) are considered to reflect a poor attentional state. To test whether an alternative characterization of RT-based attentional states yields different results, we performed a secondary analysis for each Study. We compared online task performance and subsequent memory as a function of faster RTs vs. slower RTs (from a median split) during encoding.

We found that this approach was generally inferior to the VTC analysis, in that it could not reliably detect differences in online errors based on attentional state. The strongest support for this approach came in Study 4, for which there was a significant effect for online errors (more errors for attentional states associated with faster vs. slower RTs) and a significant effect on recall (worse recall for attentional states associated with faster vs. slower RTs). Across all Studies, however, this approach — like the VTC analysis — failed to reveal differences in the temporal organization of recall based on attentional states at encoding. Thus, our choice of the VTC analysis over this alternative approach does not change our main conclusions. Nevertheless, other measures of attentional fluctuations, such as pupil diameter changes linked to physiological arousal (van den Brink et al., 2016; Unsworth et al., 2018; Clewett et al., 2020; Decker et al., 2020; W. Zhang et al., 2020), could be used in future studies to link attentional fluctuations to the temporal organization of recall.

A third possibility is that, in our Studies, attentional fluctuations had a more minor effect on recall than other variables did. For example, recall can also be structured by the semantic similarity of encoded items (Howard & Kahana, 2002b; Polyn et al., 2009; Healey et al., 2019). Individuals likely used semantic information to guide recall in our Studies, but we nevertheless

observed consistent and reliable temporal structure in recall as well: every Study showed a robust temporal contiguity effect in recall. This is in line with prior work demonstrating that, while many variables — such as list length, presentation times, incidental vs. intentional encoding, emotional salience, and inter-item distraction — can influence the the magnitude of the temporal contiguity effect, it tends to be reliably present (Healey et al., 2019; Dester et al., 2020; Lazarus et al., 2020; Peris-Yague et al., 2021).

Future studies could determine if semantic clustering differs for items encoded "in the zone" vs. "out of the zone". If so, attentional modulation of semantic but not temporal clustering would provide additional evidence for dissociations between these aspects of recall organization (Howard & Kahana, 2002b; Polyn et al., 2009; Healey et al., 2019) and provide further constraints for models of memory.

Recall tests only allow assessment of memories that are sufficiently strong as to be brought to mind without external cues. Thus, one interesting potential mechanism for the general lack of difference in overall recall between attentional states is that the items that were recalled were particularly distinctive, memorable, or salient to a given participant. This may make memory for those items less susceptible to attentional fluctuations than they otherwise would be. We used lists of relatively well-known objects to try to minimize large differences in salience or distinctiveness, but this does not rule out that some objects may be more salient or distinctive for any given participant or more memorable generally (Isola et al., 2011; Bainbridge et al., 2013; Bainbridge, 2019, 2020). However, the presence of more vs. less memorable items does not necessarily explain similar levels of temporal structure in recall across attentional states. Because items were randomly ordered for every participant, particularly memorable items were unlikely to be temporally clustered in a similar way across "in the zone" and "out of the zone" states. That

said, a compelling hypothesis is that items that are more distinctive or otherwise memorable "stick" in memory and can overcome periods of low attention to support not only their recall but also access to the surrounding temporal context, and thus recall of adjacent items. This mechanism could be explicitly tested in future work inspired by our findings. Prior relevant work has shown that item memorability and attentional states independently predict recognition memory (Wakeland-Hart et al., 2022), but such work could be extended to determine how attentional states and item memorability interact to support overall recall and the temporal structure of recall.

Moreover, other studies could use alternative measures of capturing the temporal structure of memory that do not depend on free recall (e.g., recognition or cued recall), and thus allow assessment of weaker memories. For instance, a study by Schwartz et al., (2005), used a temporally structured recognition memory task to examine the temporal organization of memory. They showed that, when individuals recognized a scene with high confidence, the probability that the next scene would also be recognized with high confidence decreased as the encoding distance between those scenes increased. Thus, future work can index attentional fluctuations during the encoding phase of such a task, and relate these fluctuations to subsequent recognition memory and its sensitivity to the temporal structure of the test. This would allow examination of how attentional fluctuations influence memories that are too weak to be recalled but can nevertheless be recognized. Indeed, other studies have found that attentional fluctuations at encoding influence recognition memory overall (deBettencourt et al., 2018; Wakeland-Hart et al., 2022), leaving open the possibility that such fluctuations also influence the temporal structure of recognition memory.

Finally, it is worth noting that our Studies deviated from standard list-learning recall tasks (Kahana, 1996; Sederberg et al., 2010; Healey & Kahana, 2014) in several ways. For example, we used images instead of words; we had longer lists of items and fewer lists, and we used gradual transitions between images (in Studies 1, 2, and 4). The changes we made were, however, necessary to address our question. Images, and gradual transitions between them, were needed to align with the traditional gradCPT approach, which was used to validate the VTC analysis and identification of "in the zone" vs. "out of the zone" states (Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014, 2015; Fortenbaugh et al., 2015; Esterman et al., 2017; Fortenbaugh et al., 2018; Rothlein et al., 2018; Esterman & Rothlein, 2019; Yamashita et al., 2021). Images are used in several studies of recall organization (Nguyen & McDaniel, 2015; Clark & Bruno, 2021; Kelly & Beran, 2021; also see Healey et al., 2019); other studies have also examined recall organization for naturalistic event memory (e.g., Moreton & Ward, 2010; Uitvlugt & Healey, 2019; Diamond & Levine, 2020). As in those studies, we replicated typical properties of lag-CRP curves in our experiments. Thus, our use of images is unlikely to have led to qualitative differences in how recall operated. We also removed the gradual transitions between images in Study 3, but still failed to find a link between attentional states and recall organization.

Our lists were longer than those typically used in list-learning studies because blocks had to be long enough to induce attentional fluctuations; attention is unlikely to robustly fluctuate in a shorter list of 10-25 items. Likewise, we had fewer lists because individuals encoded more items within each list. Despite these departures from list-learning studies, however, we were able to replicate typical properties of lag-CRP curves. Even though, in our experiments, absolute lag-CRP values at nearby lags were lower than those in other studies (e.g., Kahana, 1996; Sederberg

97

et al., 2010; Healey & Kahana, 2014), the typical shape of the lag-CRP curve, and effects of

temporal contiguity and forward asymmetry, were observed. Indeed, previous work has noted

that, although lag-CRP values are lower for longer lists, both short and long lists show clear

temporal contiguity effects (Healey et al., 2019). This highlights the reliability and robustness of

temporal organization in recall (Healey et al., 2019). Critically, as noted above, these deviations

from standard procedures were not random, but were necessary for us to test our hypotheses.

Because we were able to replicate not only typical lag-CRP findings, but also findings from the

sustained attention literature (Esterman et al., 2014, 2015; Rosenberg et al., 2015; Kucyi et al.,

2016; Esterman et al., 2017; Kucyi et al., 2017; Fortenbaugh et al., 2018; Rothlein et al., 2018;

Esterman & Rothlein, 2019; Yamashita et al., 2021), we believe it is unlikely that our design

decisions qualitatively altered attention and memory as studied in those procedures.

1.6.3 Suggestions for future research

An important contribution of this research is that it highlights the difficulty of balancing

the demands of measuring sustained attention and recall. Taxing sustained attention requires a

long, boring, and repetitive task — task features that work against robust recall for the

encountered items. Thus, the ideal task for addressing how attentional fluctuations affect the

temporal structure of recall must be long enough to induce lapses in attention but not so long that

memory performance is at floor.

One key strength of the multiple experimental designs used across our four Studies is in

building confidence in the null effect of attentional state on the temporal organization of recall:

our lists lengths varied from 30 to 120 items, our tasks varied the encoding question, and we

tested both gradual and abrupt image onsets. Thus, our findings should be useful for the field,

particularly for researchers who are concerned about potential confounds of fluctuating attention

during memory encoding and how that may affect recall organization. Future research on this topic would also benefit from exploring myriad experimental procedures to ensure that the results are not specific to any given paradigm.

Given our null effect of attentional state on recall organization across the many different experimental designs we used, we believe it is unlikely that attentional fluctuations in standard list-learning tasks reliably influence the temporal organization of recall within a list. This is because encoding lists in those tasks are likely too short to elicit robust attentional fluctuations. Nevertheless, there are multiple alternative ways that future research can assess the relationship between attentional fluctuations and memory organization, which we summarize here. We hope these strategies may help researchers address similar questions while avoiding the task complexities that we encountered.

First, our approach focused on assessing attentional fluctuations within long lists of items, in keeping with the standard approach in sustained attention research. But it is likely that attention also fluctuates slowly over the course of an entire experimental session (indeed, we found that RTs became progressively more variable across the session). Slow changes in attentional lapsing may influence recall organization from early to late in the experiment. Researchers could capitalize on such slower fluctuations by implementing the variance time course (VTC) analysis over RTs from the entire experiment rather than within a list, which has the benefit of allowing researchers to use many short lists (e.g., 15 lists of 16 items, or similar) as is often done in standard list-learning tasks (e.g., Murdock Jr., 1962; Kahana, 1996; Sederberg et al., 2010; Cortis et al., 2015; also see Healey et al., 2019). This approach would allow researchers to determine if most "in the zone" states are early on in the experimental session rather than later, and whether that is related to more temporal organization for lists encoded early

vs. late in the experiment. Alternatively, researchers could use pupil measures of attentional states (Kang et al., 2014; Unsworth et al., 2018; Unsworth & Robison, 2016; van den Brink et al., 2016; Clewett et al., 2020; Decker et al., 2020; W. Zhang et al., 2020) to measure fluctuations in attention across an experimental session, and determine if that is related to reduced temporal clustering in later lists vs. earlier lists.

Researchers could also use tests of cued recall or temporally structured recognition tests, such as the recognition approach reviewed earlier (Schwartz et al., 2005). Such tests allow assessments of temporal structure in memory while also probing weaker memories that may not be freely recalled. For example, researchers could determine the likelihood that, given an encoded item as a cue, the participant recalls the next item encoded (lag +1) as opposed to the previous encoded item (lag -1) or items encoded further away (lags > 1). These approaches overcome the limitation that only relatively strongly encoded items are freely recalled, and such items may be salient or interesting enough to "survive" periods of low attention. Such approaches could also be combined with measures of item distinctiveness or memorability (Wakeland-Hart et al., 2022; see Exploring reasons for the null effect of attentional states on recall organization), allowing researchers to determine separate and interacting effects of attentional states and memorability. For example, item memorability may increase recall likelihood more for items encountered during poor vs. good attentional states. Furthermore, more memorable items may aid in the recovery of temporal context and trigger recall of the subsequently encoded item when they are presented as a cue in a cued recall task.

Finally, the goal of our project was to examine how attentional fluctuations influence the temporal organization of recall, leaving open the question of how such fluctuations influence the semantic clustering of recall. Thus, all of the above approaches could be repeated with semantic

rather than temporal clustering as the critical output measure. Together, these approaches would yield insight into how various measures of attentional fluctuations (response time vs. pupil measures, within-list vs. across-experiment) influence different types of recall organization (temporal vs. semantic). Such work could lead to a mechanistic understanding of how various factors can overcome poor attentional states to support structured memories.

1.6.4 Relation to prior work

While we did not find evidence supporting our hypothesis that spontaneous fluctuations in attention influence the temporal organization of memory, it is likely that stronger manipulations of attention would affect recall organization. Indeed, there is ample work showing that attention influences memory (see Chun & Turk-Browne, 2007; Aly & Turk-Browne, 2017). Studies involving dual tasks show that divided attention at encoding is associated with worse memory at retrieval (for e.g. Baddeley et al., 1984; Craik et al., 1996; Uncapher & Rugg, 2005). Other tasks involving experimenter-manipulated attention also show robust effects on memory (for example, Yi & Chun, 2005; Uncapher & Rugg, 2009; Aly & Turk-Browne, 2016). One recent study showed that the type of attention at encoding (whether there was a semantic task or no task) influences temporal clustering in recall (Long & Kahana, 2017). Another recent study showed that introducing distractions during encoding disrupts the temporal contiguity effect (Cutler et al., 2020). Thus, experimentally manipulated attention robustly impacts memory performance generally as well as the temporal organization of recall.

There is evidence that spontaneous attentional fluctuations influence memory, but these studies do not examine the temporal dynamics of recall. One such body of work is research on mind wandering. In these studies, participants are asked to report whether they are "on-task" or "off-task" at various intervals (for example, Smallwood et al., 2003, 2008; Metcalfe & Xu, 2016;

Xu & Metcalfe, 2016; Xu et al., 2018; also see Smallwood & Schooler, 2015; Christoff et al., 2016). These studies have found that more mind wandering is associated with decreased change detection in memory (Garlitch & Wahlheim, 2020), less precise cued recall (Martarelli & Ovalle-Fresa, 2021), and worse memory performance in general (Smallwood et al., 2003; Risko et al., 2012). Furthermore, as mentioned above, there is evidence that worse attentional states at encoding, as indexed by RTs, are associated with worse recognition memory (deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022) and worse associative memory (Elshiekh & Rajah, 2021) in a subsequent test.

Together, the above-reviewed research shows that experimental manipulations of attention have effects on the temporal structure of recall and that spontaneous attentional fluctuations have effects on other forms of memory. Our current findings suggest that temporal context may be a more powerful driver of memory organization than attentional fluctuations — at least those that occur naturally and spontaneously during a task. Determining the conditions in which spontaneous attentional fluctuations may influence the temporal structure of memory requires further work; for example, future studies using temporally structured recognition tests or alternative measures of attentional fluctuations (e.g., pupil diameter) may yield important insights.

1.6.5 Conclusion

Across four Studies, we did not find any evidence that attentional fluctuations during encoding, as measured by the speed or variability of response times, influenced the temporal organization of recall. Temporal organization of recall is robust, even for memories encoded during relatively poor attentional states. These findings suggest that temporal context serves as a strong scaffold for episodic memory, one that can overcome spontaneous fluctuations in

attentional states. Future research can assess the generality of these results by examining other measures of attention and memory.

Chapter 2: Pupil size at encoding predicts the temporal structure of memory

Manasi Jayakumar, Aria Tsegai-Moore, Cecilia Nemeth, & Mariam Aly

2.1 Introduction

We remember events in a temporally organized manner: recalling a specific event acts as a cue for recall of other events that occurred close to it in time (Howard & Kahana, 2002a). Temporal organization of memory is extremely robust (Healey & Kahana, 2014; Healey et al., 2019): it is observed in both tightly controlled laboratory experiments (Kahana, 1996; Healey et al., 2019; Jayakumar et al., 2023) and naturalistic settings (Moreton & Ward, 2010; Diamond & Levine, 2020), for various types of materials (Cortis et al., 2015; Nguyen & McDaniel, 2015; Healey et al., 2019), across various age group (Healey et al., 2019; Diamond & Levine, 2020), and is even evident in eye movement behavior (Kragel & Voss, 2021). Temporal organization can be modulated by many factors, such as the list length or presentation speed (Healey et al., 2019); but it is an extremely robust hallmark of memory in healthy adults. Yet, most prior studies have failed to measure how the temporal structure of subsequent memory might be affected by spontaneous fluctuations in attention during learning, as regularly happens in daily life when we "zone out" during a mudane experience (Smallwood & Schooler, 2006, 2015). Here, we test whether the strength of the temporal organization of memory is modulated by natural fluctuations in attention. We hypothesize that attention is a key glue that binds together items experienced nearby in time, such that disruptions in sustained attention would weaken the temporal structure of memory.

We initially tested this hypothesis in a series of behavioral studies, using a modified sustained attention task followed by a free recall memory test (Jayakumar et al., 2023). Surprisingly, across four studies, we found that fluctuations in attentional states, as defined by variability in response times, did not predict differences in the temporal organization of recall. Instead, the hallmarks of temporally organized recall – the tendency to recall items together if

they were encoded closer vs. farther together in time, and the tendency to recall items in a forward rather than backward direction – were entirely preserved even when encoding occurred in a worse attentional state. One possible explanation for these results is that temporal context representations that group nearby experiences are extremely robust, able to overcome periods of poor attention. An alternative hypothesis is that response-time-based measures of attentional fluctuations do not capture the kinds of cognitive changes that predict temporal binding in memory. Here, we test these competing hypotheses by using both pupil-based measures and response-time-based measures of attentional states and examine their ability to predict the temporal organization of subsequent recall.

Response times and pupil size have both been extensively used in prior studies of sustained attention. Response time variability (e.g., Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014; Rosenberg et al., 2017; Fortenbaugh et al., 2018; Esterman & Rothlein, 2019) and speed (e.g., Robertson et al., 1997; Manly et al., 1999; Cheyne et al., 2006; deBettencourt et al., 2018; Decker et al., 2022; Wakeland-Hart et al., 2022) robustly predict both online task performance (e.g., Esterman et al., 2013; Rosenberg et al., 2013; Esterman et al., 2014; Rosenberg et al., 2017; Fortenbaugh et al., 2018; Esterman & Rothlein, 2019) and subsequent recognition memory (e.g., deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022). In addition to response time measures, pupil size has often been used as an indicator of attention and arousal, with many studies observing consistent effects or coupling across these measures (e.g., Nuthmann & Van Der Meer, 2005; Reinhard et al., 2007; Seymour et al., 2013; Unsworth & Robison, 2016; Wainstein et al., 2017; Kozunova et al., 2022; Burlingham et al., 2022).

The use of pupil size to index fluctuations in attention and arousal is motivated by animal models and computational modeling approaches demonstrating that pupil size can track activity of the Locus Coeruleus (LC) and its release of the neurotransmitter norepinephrine (Aston-Jones & Cohen, 2005); in particular, low-arousal states have been associated with reduced baseline pupil diameter (Aston-Jones & Cohen, 2005; Megemont et al., 2022). Pupil size has also been linked to activity of the basal forebrain cholinergic system (Joshi & Gold, 2020), a system associated with externally oriented attention and memory encoding (Newman et al., 2012; Tarder-Stoll et al., 2020; Poskanzer & Aly, 2023; also see Ruiz et al., 2021). Investigations of human behavior have shown that pupil size is tightly linked to cognitive load, arousal, attentional performance, and subsequent memory (Kahneman & Beatty, 1966, 1967; Kafkas & Montaldi, 2011; Otero et al., 2011; Papesh et al., 2012; Rijn et al., 2012; Naber et al., 2013; Ariel & Castel, 2014; Hoffing & Seitz, 2014; van den Brink et al., 2016; Unsworth et al., 2018, 2020; Unsworth & Miller, 2021; Whitlock et al., 2023). Larger pupil size is associated with on-task compared to mind-wandering periods (Grandchamp et al., 2014). For example, baseline pupil diameter, i.e., pupil size before trial onset, has been successfully used as an indicator of attentional fluctuations and a predictor of behavioral performance (Gilzenrat et al., 2010; Unsworth & Robison, 2016; van den Brink et al., 2016; Unsworth et al., 2018).

In the current study, we used both pupil size and response times to index attentional fluctuations during the encoding of object images, and then tested how these measures relate to both online task performance and subsequent memory in a free recall task. Using both measures allowed us to examine whether the temporal organization of recall is robust to attentional fluctuations no matter how they are measured, or if the two measures of attention diverge in the aspects of behavior that they predict.

Participants were presented with a series of trial-unique object images, making a judgment on each with concurrent eye-tracking. This encoding was followed by free recall, in which participants recalled the objects verbally in any order they chose. We characterized attentional states based on a median split of pre-trial pupil sizes at encoding: larger pupil sizes indexed "better" attentional states, while smaller pupil sizes indexed "worse" attentional states. Secondarily, we defined attentional states based on response times (RTs), with faster RTs indicating "worse" attentional states and slower RTs indexing "better" states (deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022). We then tested how online task performance, recall rates, and the temporal organization of recall varied by attentional state at encoding. To foreshadow our results, we found that pupil-based and response-time-based measures of attention predicted different aspects of behavior, with only pupil size robustly predicting the temporal organization of subsequent memory.

2.2 Methods

2.2.1 Inclusion criteria

Participants in our final sample (n = 71) had to meet a number of inclusion criteria. First, individuals had to report normal or corrected-to-normal vision with contact lenses or LASIK surgery. We opted to not include any participants with eyeglasses because glasses can interfere with the eye-tracking calibration and validation procedures (Kang et al., 2014). Second, we only included individuals for whom we had sufficient eye tracking data, after data loss due to calibration issues, blinking, off-screen fixations, etc. To be included, a participant needed to have eye tracking data for at least 50% of time points within a trial (Clewett et al., 2020; also see Winn et al., 2018) for 97% of trials in at least two (out of three) blocks. Finally, we only

included individuals who adequately performed the behavioral task during the encoding phase (at least 80% accuracy, following Jayakumar et al., 2023).

2.2.2 Participants

We determined our minimum sample size to be 50 participants based on a priori power analyses for 80% power at an alpha of 0.05 in a within-participant design. As in our behavioral work (Jayakumar et al., 2023), this power analysis was based on estimated effect sizes for 1) temporally organized recall, particularly main effects and interactions of absolute lag and direction (Kahana et al., 2002; Spillers & Unsworth, 2011; Palombo et al., 2019; Diamond & Levine, 2020; Jayakumar et al., 2023); and 2) interactions between absolute lag/direction and other variables (Palombo et al., 2019; Diamond & Levine, 2020). We opted to exceed the minimum sample size from the power analysis to overcome effect size overestimation (Brand et al., 2008; Bakker et al., 2012). Our final sample size was therefore 71 participants. Of these, 64 participants were recruited from the Columbia University undergraduate participant pool in exchange for course credit, and 7 participants were recruited through flyers around the Columbia University community and were compensated with cash ($15/hour). Participants were between 18-35 years of age and provided informed consent as per the Columbia University Institutional Review Board.

In terms of gender, 40 participants identified as female, 30 participants as male, and 1 participant as non-binary. In terms of race, 32 participants identified as White, 5 as Black or African American, 21 as Asian, 2 as Hispanic/Latino, 2 as Middle-Eastern or Arab, and 9 as multiracial (White and Black (N = 4), White and Asian (N = 1), Asian and Black (N = 1), White and American Indian/Alaska Native (N = 2), White and Middle-Eastern (N = 1)). In terms of ethnicity, 10 participants identified as Hispanic/Latino, and 61 participants identified as not

Hispanic or Latino. Participants ranged in age from 18-34 (mean age = 20.70, SD = 4.23) and in education from 12-19 years (mean education = 13.73, SD = 1.60).

2.2.3 Stimuli

We chose 355 images of common objects from pre-curated object databases pre-curated object databases, including stimuli from the Mnemonic Similarity Task (Yassa et al., 2011; https://faculty.sites.uci.edu/starklab/mnemonic-similarity-task-mst), the ecological alternative to the Snodgrass and Vanderwart stimulus set (Moreno-Martínez & Montoro, 2012), the Interaction Envelope (Bainbridge & Oliva, 2015a, 2015b; http://www.wilmabainbridge.com/datasets.html), the DinoLab Object Database (Davis et al., 2020; https://mariamh.shinyapps.io/dinolabobjects), the MultiPic (Duñabeitia et al., 2018; https://www.bcbl.eu/databases/multipic/), and the Bank of Standardized Stimuli (Brodeur et al., 2014; https://sites.google.com/site/bosstimuli).The images were converted to .jpg (if they were in a different format) and resized to a 500-pixel square. To prevent luminance-driven changes in pupil size, we created isoluminant images using the script and procedure described in Zhang & Emberson, 2019. However, creating these isoluminant images sometimes distorted the object and made it difficult to identify. We thus ran a small pilot study with an object-naming task, in which 16 participants were presented with isoluminant images of objects and asked to identify the object in the image. For our main experiment, we selected 255 images that were identifiable by at least 70% of the participants who saw that image in the pilot study.

2.2.4 Experimental Design and Procedure

The experiment was conducted in 3 blocks, each of which included a study, distractor, and recall phase (Fig 15). Eye-tracking data were collected during the study phase of each block (see Pupil Data Acquisition)

During each study phase, we presented participants with 80 trial-unique isoluminant

images of common objects (see Stimuli above). Each image was presented on the screen for
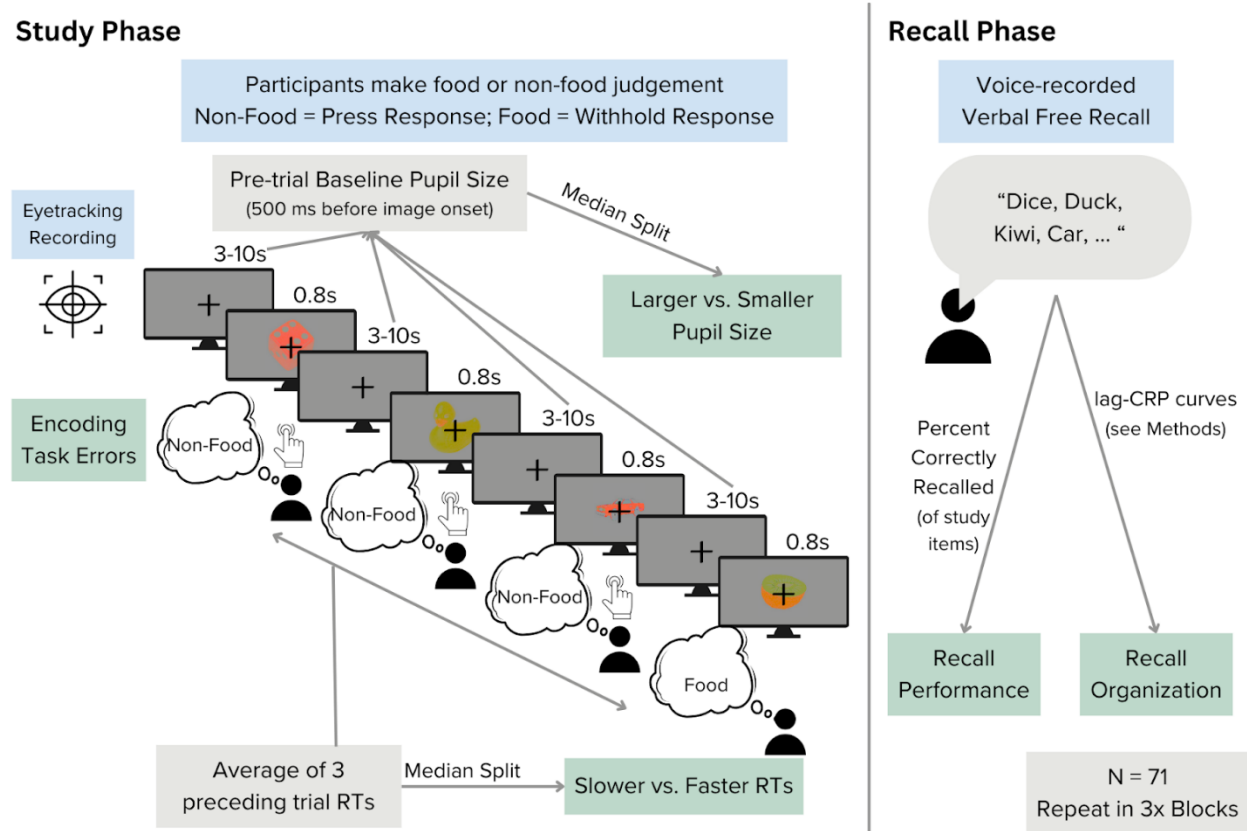
Figure 15. Schematic of the experimental design and task measures. Participants completed the task in three blocks, each of which contained three phases: study phase with concurrent eye-tracking, distractor phase (not shown), and recall phase. During the study phase, participants were presented with a series of isoluminant images of objects (see Stimuli). For each image, they were asked to press a button for a non-food item and withhold the button press for a food item. Attentional states were indexed using both the pre-trial baseline pupil size (median split of the residualized pupil size in the 500ms prior to stimulus onset) and response time (median split of the average of the three preceding trial RTs). Following a math distractor phase (not shown in fig.), participants were asked to verbally recall objects from the study phase in any order they choose. Recall performance (percent correctly recalled of study items) and temporal organization (using lag-CRP curves, see section 2.2.6.6.) were measured from the recall phase.

800ms with a fixation cross in the center of the image. The light intensity across the screen was

kept uniform and isoluminant to the image. Throughout the study block, participants were asked

to fixate on the cross in the center of the screen (including before, during, and after image

111

presentation). The inter-trial interval (ITI) – the duration between two images – was equally distributed between 3 to 10 seconds in 500ms increments (Unsworth et al., 2018). These ITIs were chosen based on past work showing that long and unpredictable ITIs are more likely to lead to attentional lapses than short, predictable ITIs (Unsworth et al., 2020).

In the study phase, participants performed a go/no-go task on the object images. This task was chosen because it is commonly used in sustained attention research, including our prior work (Jayakumar et al., 2023; see Esterman & Rothlein, 2019 for a review). For each presented image, participants were asked to judge if the object depicted by the image was a food or non-food item (Jayakumar et al., 2023). If the object was not a food item, participants were asked to press a button ("go" trials). If the object was a food item, participants were asked to withhold their response ("no-go" trials). The non-food trials occurred 90% of the time, with the food trials occurring only 10% of the time. Thus, in each block of 80 images, 72 images were non-food, while 8 images were food items. Because the "go" response is the dominant response, individuals are likely to experience occasional attentional lapses (Esterman & Rothlein, 2019) , indicated by fast, habitual responses that predict errors..

Following each study phase, participants performed a distractor task (70 seconds) to discourage any rehearsal in the delay between memory encoding and subsequent recall. During this distractor phase, participants were presented with 2-digit additions and subtractions. Participants had to judge whether the solutions presented with the math problem were correct or incorrect using one of two keys.

The final phase in each block was the free recall task. Here, participants were asked to verbally recall as many objects as they could in any order they chose from the corresponding study phase. Participants were given 2.5 minutes to perform this recall, with a countdown timer

on the screen indicating the time remaining. The 2.5-minute verbal recall was recorded with a microphone.

Before beginning the three experimental blocks, participants performed the task in a practice block. The practice block was identical to the experimental blocks, except that 15 images were presented during the study phase, of which 4 images were food items. The data from the practice block were excluded from further analyses.

2.2.5 Pupil Data Acquisition

An Eyelink 1000 Plus eye-tracker by SR Research was used during the study phase to record pupil size. We tracked participants' right eye at 1000Hz, using a head-stabilized mount. Stimuli were displayed using a 1920px x 1080px resolution, 24-inch LED monitor with a refresh rate of 60 Hz. Participants were seated 93cm away from the computer. They were instructed to remain as still as possible during the study phase and fixate on the cross in the center of the screen. Before beginning each study phase, participants performed a 9-point eye tracking calibration and validation procedure.

2.2.6 Analyses

2.2.6.1 Preprocessing of eye tracking data

The Eyelink 1000 Plus stores raw eye-tracking and pupil data in an.EDF format. Using the EDF2ASC converter provided by SR Research, we converted the EDF files to asc text files. We then used the 'eyelinker' R package (https://cran.r-project.org/package=eyelinker) to read the asc files. First, we extended the eye-tracker detected blinks by 100ms on either side to remove artifacts (Winn et al., 2018; Mathôt & Vilotijević, 2023). Second, to correct for the change in apparent pupil size when participants' eyes move away from fixation, we performed a pupil foreshortening error correction. Following the geometric model procedure introduced by Hayes

& Petrov (2016), we estimated the pupil foreshortening error as the cosine of the angle between the eye-to-camera and eye-to-stimulus axis. We then divided the raw pupil size by the square root of the pupil foreshortening error to obtain the corrected pupil size (Hayes & Petrov, 2016). Third, we interpolated for blinks and off-screen fixations using the 'na.approx()' function in 'zoo' R package (https://zoo.r-forge.r-project.org/). Finally, to correct for time-dependent linear drift in pupil size, we regressed out trial numbers from pupil size within participants and extracted residuals (Decker et al., 2020). The residual pupil size data were used for subsequent analyses.

2.2.6.2 Pupil measures of attention

We focused our analysis on pre-trial measures of pupil size, based on prior work linking these pre-stimulus pupil measures to attentional fluctuations (van den Brink et al., 2016; Unsworth et al., 2020). Using the pre-processed and residualized pupil size, we calculated the average pupil size preceding each trial as the mean of the residual pupil size in the 500ms window before image onset, during which a fixation cross was presented. We then performed a median split of these mean pupil size values to create two conditions: larger vs. smaller pupil diameter. Following past work, we considered larger pupil diameters to correspond to better attentional states (van den Brink et al., 2016; Unsworth et al., 2020).

2.2.6.3 Secondary measures of attention

Our primary measure of interest for indexing attentional states at encoding was pre-trial pupil diameter, but we also examined secondary measures of attention. Response-time-based measures of attention reliably predict online task performance (e.g. Robertson et al., 1997; deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022; Jayakumar et al., 2023). For example, faster response times (RTs) in a go/no-go task indicate habitual responding

114

and are more likely to precede errors than slower RTs (Rosenberg et al., 2013). In our prior work, however, we found that response-time-based measures of attention do not predict the temporal structure of recall (Jayakumar et al., 2023). To replicate and extend these findings, we used RTs during the go/no-go task as a secondary index of attentional state. We defined the attentional state during the encoding of an object based on the response time of the three trials preceding that object (following deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022). We focused on RTs preceding a given trial both because prior studies have found that RTs to the three preceding trials predict attentional lapses on a given trial (deBettencourt et al., 2018; Decker et al., 2020; Wakeland-Hart et al., 2022) and because focusing on pre-trial measures aligns the RT approach to the pre-trial measures of pupil diameter that were of primary interest.

To calculate RT-based attentional states based on the preceding three trials, we first controlled for linear drift throughout the task by removing the effect of trial number on RTs and examining the residuals. Second, for every trial, we computed a rolling mean over the residual RTs from the three prior trials using the function *rollapply()* from the R package zoo (https://zoo.r-forge.r-project.org/). We then conducted a median split over these values to sort trials into "better" (slower preceding RTs) vs. "worse" (faster preceding RTs) attentional states.

2.2.6.4 Errors at encoding

We examined whether there was a difference between the two attentional states in the errors participants made during the encoding task. Errors were calculated as the percentage of incorrect responses for the study items, including incorrect button presses to the "no-go" food trials and incorrectly withheld responses to the "go" non-food trials. A paired-samples t-test was conducted to examine whether errors differed across trials with smaller pupil size (worse

attentional state) vs. larger pupil size (better attentional state). This analysis was also repeated

with the secondary measures of attention to determine whether errors differed for trials with

worse vs. better attentional states as defined by RTs (see Secondary measures of attention). We

conducted group-level analyses using a paired-samples t-test.

2.2.6.5 Recall performance

We computed overall recall performance as the percentage of items that were correctly

recalled across all blocks. We also calculated recall performance for each attentional state as the

percentage of items encoded in that attentional state that were subsequently correctly recalled.

Because the attentional states were defined by a median split of pupil diameter or RT measures,

the number of items encoded in each state was the same. We conducted group-level analyses

using a paired-samples t-test.

2.2.6.6 Temporal organization of recall

We measured the temporal organization of recall using lag-Conditional Response

Probability (lag-CRP) curves (Kahana, 1996). The lag-CRP curve quantifies the probability of

recalling two items successively based on their relative positions, or lag, at encoding. To plot

these curves, we first examined all pairs of successively recalled items, and calculated the lag of

the recall transitions as the difference in those items' position numbers at encoding. This lag can

be positive or negative, depending on whether the second item was encoded after (positive) or

before (negative) the first item. We then summed the number of recall transitions that occurred at

each lag, giving the observed number of recall transitions. This observed number of recall

transitions was then divided by the number of opportunities to make a recall transition at that lag,

which yields the Conditional Response Probability (CRP) at that lag (Kahana, 1996). As in our

prior study and as is standard in lag-CRP approaches, transitions to and from repetitions and

intrusions (items recalled from a prior study list) were excluded from this analysis (Kahana, 1996; Jayakumar et al., 2023).

Lag-CRP curves depict two characteristic features of temporal organization: temporal contiguity and forward asymmetry (Kahana, 1996). Temporal contiguity refers to the higher likelihood of recalling items closer together in time if they were also encoded closer together (Kahana, 1996). Temporal contiguity can be seen in the lag-CRP curve as the peak in probability at ±1 lag, followed by decreasing probabilities at larger lags. Forward asymmetry refers to the higher likelihood of recalling items in the forward direction (i.e. in the direction of encoding) vs. backward direction. This can be seen in the lag-CRP curve as the larger conditional response probability for positive lags vs. negative lags. Finally, these effects are interactive, such that the tendency to recall items in the forward vs. backward direction is enhanced for nearby lags.

Our primary hypothesis was that better vs. worse attentional states, as indexed by pupil diameter, are more conducive to maintaining a temporal context representation that would facilitate temporally organized recall. Thus, we predicted that temporal contiguity and/or forward asymmetry of recall should be stronger for items encoded in better vs. worse attentional states. To test this, we constructed separate lag-CRP curves for each attentional state, following the procedures of our prior study (Jayakumar et al., 2023). First, pairs of successively recalled items were labeled as being in the "better" attentional state if they were both encoded in that state, or labeled as being in the "worse" attentional state if they were both encoded in that state. If the two successively recalled items were encoded in different attentional states, we labeled those items as a transition pair; these transition pairs were excluded from our lag-CRP analyses because our main hypotheses were about recall within each state separately.

To plot the lag-CRP curve separately for each attentional state ("better" and "worse" states), we only considered successively recalled items that were both labeled as being from that particular state. Observed recall transitions were calculated as before – the lag between each pair of successively recalled items, as defined based on their encoding positions. We then divided these observed recall transitions at each lag by the number of times a recall transition at that lag could have occurred, regardless of attentional state. This yields the Conditional Response Probability (CRP) for each lag within each attentional state. The CRPs for each lag were calculated within-block and then averaged across blocks resulting in a single CRP value per participant at each lag for each attentional state.

To test for a statistical difference in the temporal organization of recall between the two attentional states, we performed a three-way repeated-measures ANOVA with attentional state (better vs. worse), absolute lag (1 to 29), and recall direction (forward vs. backward) as factors. The sphericity assumption was tested using Mauchly's test of sphericity and the Greenhouse-Geisser correction was applied when the assumption was violated.

2.3 Results

2.3.1 Pupil diameter as an assay of attentional state at encoding

During the study phase, participants viewed images of objects and judged each as being a food or a non-food item (Fig. 15). Pupil diameter measures were obtained during the study phase using an eye-tracker and pre-processed as described in Methods: Analyses.

We first examined whether pupil size declines over time within a block, demonstrating the time-on-task effect (e.g. van den Brink et al., 2016). A linear regression analysis revealed that pre-trial pupil diameter (as defined in Methods: Analyses) significantly declined over trials within a block ($F_{1,14958} = 285.1$, $p < 0.0001$, adj. $R^2 = 0.019$). This provides a preliminary

indication that pupil diameter is sensitive to attention and/or arousal in our task. Because we were primarily interested in moment-by-moment changes in attention rather than slower changes over a block, we proceeded to extract residuals by regressing out trial numbers from pupil size. For each trial, we then averaged the residual pupil size over the 500ms window before stimulus onset to obtain the pre-trial baseline pupil size. This was followed by a median split of this measure (see Fig. 15) to sort trials into those associated with a larger vs. smaller pre-trial pupil size. Based on prior research, larger pupil diameters were considered to indicate a better attentional state (van den Brink et al., 2016; Unsworth et al., 2020).

Before examining our main hypothesis about the temporal organization of recall, we first examined whether pre-trial pupil size predicted accuracy in the go/no-go task and overall recall performance. This allowed us to determine whether pupil size predicted all aspects of online performance and memory or if it was more selectively tied to one component of performance.

2.3.2 Pupil size does not predict online task errors

We tested whether differences between the attentional states were reflected in online performance by examining accuracy in the go/no-go task during encoding. Participants typically performed well on the food or non-food encoding judgment. Mean percentage of errors (as defined in Methods) was 3.37% (SD = 3.05%, median = 2.5%). However, there was no significant difference in task errors between the better (mean = 3.42%, SD = 3.34%, median = 2.52%) vs. worse attentional state, as defined by pupil size (mean = 3.35%, SD = 3.37%, median = 2.48%; t_{70} = 0.22, p = 0.83, Cohen's dz = 0.026, 95% CI [-0.55,0.69]; Fig. 16A).

To examine if any differences in online performance were masked by the habitual, and thus highly accurate response pattern in the frequent non-food trials, we examined performance for the infrequent food trials alone (deBettencourt et al., 2018; Wakeland-Hart et al., 2022).

Here, too, there was no significant difference between attentional states in errors on infrequent trials (t_{70}=0.41, p = 0.68, Cohen's dz = 0.049, 95% CI [-0.38,0.58]). Thus, pupil size at encoding did not predict online performance in the food /non-food judgment task.
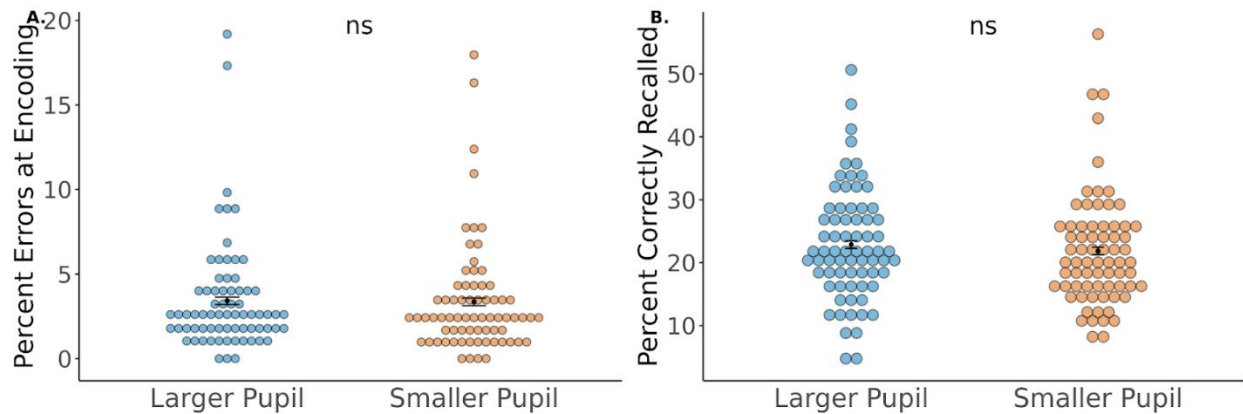
Figure 16. Encoding task performance and recall performance by pupil-based attentional states. Encoding errors and recall performance do not differ between the pupil-based attentional states. A) Individual points show the percentage of encoding judgment errors made by each participant during the "better" (larger pupil) and "worse" (smaller pupil) attentional states. There was no significant difference in the online task performance between the two attentional states. B) Individual points show the percentage of items correctly recalled by each participant as a function of whether they were encoded during "better" (larger pupil) or "worse" (smaller pupil) attentional states. There was no difference in recall performance between the two states. Black points in panels B & C indicate the mean of the measure; errors bars indicate the standard error of the within-participant difference between the "better" and "worse" attentional states. ns = not statistically significant.

2.3.3 Recall performance does not differ by pupil size at encoding

We next tested whether pupil size at encoding predicted overall recall performance. Mean recall accuracy (defined as the percentage of items encoded that were subsequently recalled) was 22.28% (SD = 8.18%).

We hypothesized that recall performance would differ between the two attentional states, such that the "better" attentional state would be associated with greater recall performance. However, there was no significant difference in recall performance between the "better" (Mean =

22.80%, SD = 8.89) vs. "worse" attentional state (Mean = 21.79%, SD = 8.68; t_{70} = 1.02, p = 0.31, Cohen's dz = 0.12, 95% CI [-0.63, 1.95]; Figure 16B). Hence, pupil size at encoding did not predict the percentage of items recalled overall.

2.3.4 Temporal organization of recall differs by pupil size at encoding

Our primary hypothesis was that "better" attentional states (as indexed by larger pupil sizes) are more conducive to maintaining temporal context representations that facilitate temporally organized recall. We predicted that this would result in stronger temporal contiguity and/or forward asymmetry of recall for "better" vs. "worse" attentional states. To examine differences in the temporal organization of recall, based on attentional state at encoding, we plotted lag-CRP curves. Figure 17A shows the overall lag-CRP curve, across blocks and participants; Figure 17B shows the lag-CRP curves across blocks and participants, separately for items encoded in the "better" (larger pre-trial pupil size) vs. "worse" (smaller pre-trial pupil size) attentional states.

We conducted a three-way repeated-measures ANOVA on the lag-CRP measures (from Fig. 17B) with attentional state, absolute lag (1 to 29), and direction (forward vs. backward) as factors. We expected to replicate prior studies (Kahana, 1996; Healey et al., 2019) showing temporally organized recall, with recall being more likely to transition to items encoded nearby vs. far away (main effect of absolute lag), and more likely to transition in the forward vs. backward direction (main effect of direction), especially for nearby lags (interaction between absolute lag and direction). Our main hypothesis was that this hallmark of temporally organized recall would differ as a function of attentional state at encoding, leading to a three-way interaction between attentional state, lag, and direction.
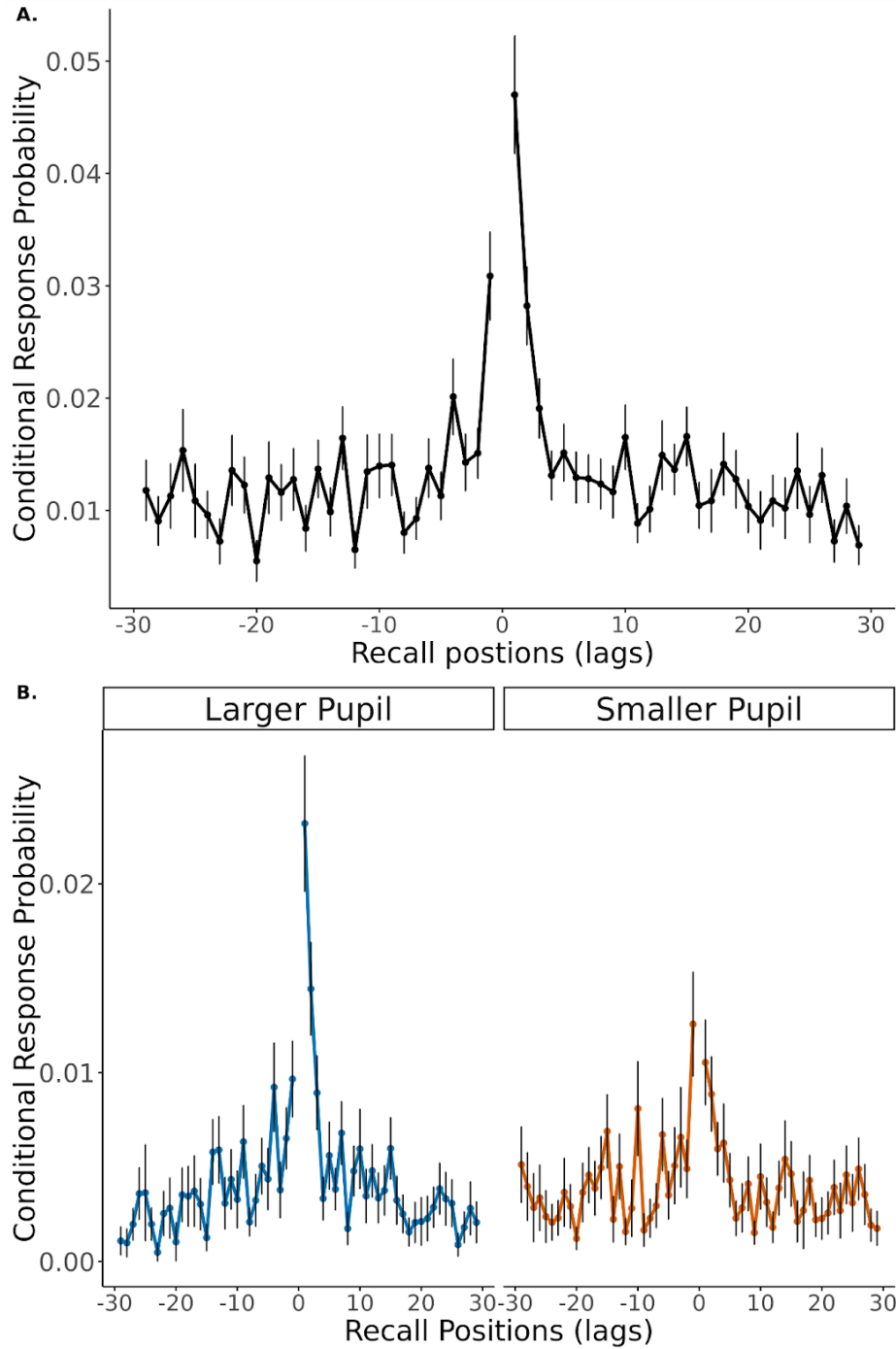
Figure 17. Lag-CRP curves overall and by pupil-based attentional state. A. Overall lag-CRP curve across participants and blocks. B. Lag-CRP curves plotted separately for items encoded in the better (larger pupil; left) and worse (smaller pupil; right) attentional state. There was a significant difference in the temporal organization between the two attentional states. Error bars represent the standard error.

Replicating prior work, first, we found a significant main effect of absolute lag ($F_{15.51,1085.39} = 9.43$, $p < 0.0001$, $\eta_p^2 = 0.12$) indicating that there was a greater likelihood of recalling items together when they were encoded closer together vs. farther away. We did not find a significant main effect of direction ($F_{1,70} = 1.43$, $p = 0.24$, $\eta_p^2 = 0.02$). However, we found a significant interaction between absolute lag and direction ($F_{16.88,1181.34} = 1.69$, $p = 0.04$, $\eta_p^2 = 0.02$), indicating that the tendency to recall items in the forward rather than the backward direction was stronger for closer vs. farther lags.

We then turned to the effects of attentional state. We did not find a main effect of attentional state ($F_{1,70} = 0.018$, $p = 0.89$, $\eta_p^2 = 0.0003$). We also did not find interactions between attentional state and absolute lag ($F_{15.59,1091.18} = 1.29$, $p = 0.20$, $\eta_p^2 = 0.02$), nor between attentional state and direction ($F_{1,70} = 2.98$, $p = 0.089$, $\eta_p^2 = 0.04$). Critically, however, our hypothesis was confirmed by a three-way interaction between attentional state, absolute lag, and direction ($F_{15.95,1116.64} = 1.95$, $p = 0.014$, $\eta_p^2 = 0.027$).

To understand the underlying pattern of results contributing to this three-way interaction, we conducted separate two-way repeated-measures ANOVAs on the lag-CRP measures within each attentional state, with absolute lag (1 to 29) and direction (forward vs. backward) as factors. For the "better" attentional state, as indexed by larger pupil sizes, we observed a significant main effect of absolute lag ($F_{15.05,1053.67} = 8.11$, $p < 0.0001$, $\eta_p^2 = 0.10$), a significant main effect of direction ($F_{1,70} = 4.47$, $p = 0.04$, $\eta_p^2 = 0.06$), and a significant interaction between absolute lag and direction ($F_{13.66,956.20} = 2.83$, $p < 0.0001$, $\eta_p^2 = 0.04$). Thus, the hallmarks of temporally organized recall were observed for "better" attentional states at encoding: these attentional states were associated both with temporal contiguity effects (main effect of absolute lag) and the

tendency to recall items more often in the forward vs. backward direction for nearby lags (interaction between absolute lag and direction).

In contrast, for the "worse" attentional state, characterized by smaller pupil sizes, we only found a main effect of absolute lag ($F_{14.87,1040.87} = 3.13$, $p < 0.0001$, $\eta_p^2 = 0.04$). There was no main effect of direction ($F_{1,70} = 0.33$, $p = 0.57$, $\eta_p^2 = 0.005$), nor an interaction between absolute lag and direction ($F_{17.43,1220.29} = 0.85$, $p = 0.64$, $\eta_p^2 = 0.012$).

We detected a main effect of lag for both "worse" and "better" attentional states, but there is a noticeable difference in the lag-CRPs for the two states at the closest lags (Fig 17B). We therefore conducted a supplementary analysis to test whether nearby recall transitions were more likely for the "better" vs. "worse" attentional state. We divided recall transitions into nearby (lags 1-14) vs. far away (lags 15-29), and compared the probability of these recall transitions based on attentional state at encoding using a repeated-measures ANOVA with attentional state (better vs. worse) and recall distance (nearby vs. far away) as factors. Here, we found a main effect of recall distance ($F_{1,70} = 41.18$, $p < 0.0001$, $\eta_p^2 = 0.37$), again reflecting the tendency to make more nearby vs far-away recall transitions. There was no main effect of attentional state ($F_{1,70} = 0.05$, $p = 0.82$, $\eta_p^2 = 0.0007$). Critically, however, there was an interaction effect between recall distance and attentional state ($F_{1,70} = 8.68$, $p = 0.004$, $\eta_p^2 = 0.11$), reflecting the tendency for nearby recall transitions to be more likely for "better" (probability of 0.0058) vs. "worse" (probability of 0.0047) attentional states, and far-away recall transitions to be more likely for "worse" (probability of 0.0034) vs. "better" (probability of 0.0025) attentional states.

To further support our findings, we conducted an additional analysis focused on the closest recall transitions. Past studies have found that forward asymmetry in recall (forward >

backward transitions) is particularly pronounced at the closest lags of ±1, such that +1 transitions are more likely than -1 transitions (e.g., Diamond & Levine, 2020; Jayakumar et al., 2023). We therefore examined forward asymmetry differences between the two attentional states at the nearby lags of ±1. A two-way repeated-measures ANOVA with lag (-1 and +1) and attentional state ("better" vs. "worse") as factors revealed a main effect of lag ($F_{1,70} = 4.57$, $p = 0.036$, $\eta_p^2 = 0.06$) and an interaction between lag and attentional state ($F_{1,70} = 6.53$, $p = 0.013$, $\eta_p^2 = 0.085$). This was driven by a significantly greater likelihood of +1 vs. -1 transitions for items encoded in the "better" attentional state ($t_{70} = 3.01$, $p = 0.004$, Cohen's dz = 0.36; 95% CI [0.004, 0.022]) but no difference in +1 vs. -1 transitions for items encoded in the "worse" attentional state ($t_{70} = 0.58$, $p = 0.56$, Cohen's dz = 0.07; 95% CI[-0.005,0.009]). Thus, there was a significant difference in forward asymmetry based on attentional state, such that encoding in a "better" (vs. "worse") attentional state was associated with a greater likelihood of recalling in the forward vs. backward direction for the ±1 recall transitions.

Overall, these lag-CRP findings suggest that a "better" attentional state at encoding (defined by larger pupil sizes) predicts stronger temporal organization of subsequent recall. "Worse" attentional states at encoding produce disruptions in temporal clustering, forward asymmetry at the ±1 lags, and forward asymmetry differences for nearby vs. farther lags as a whole. In sum, pupil size at encoding did not predict online performance or the overall amount of memory recall but did predict the temporal organization of recall.

2.3.5 Secondary measures of attention do not predict temporal organization of recall

In a prior set of studies, we found that response-time-based measures of attention did not predict the temporal organization of recall, even though they reliably indexed online task performance (Jayakumar et al., 2023). To examine if there was a similar pattern of results in the

current study, we categorized trials into "better" vs. "worse" attentional states based on RT for the go/no-go task (see Methods). For each item, we found the average response times from the three preceding trials (following deBettencourt et al., 2018; Wakeland-Hart et al., 2022). We then performed a median split of these values to define two attentional states. Trials with slower preceding response times were categorized as being in a "better" attentional state, whereas trials with faster preceding response times were categorized as being in a "worse" attentional state (deBettencourt et al., 2018; Wakeland-Hart et al., 2022).

We first examined whether online performance during the encoding task was different between the "better" vs. "worse" attentional state, as defined by preceding RTs. In line with prior work (deBettencourt et al., 2018; Wakeland-Hart et al., 2022; Jayakumar et al., 2023), we expected that the "worse" attentional state would be associated with more encoding task errors. Indeed, participants made significantly more food vs. non-food judgment errors during "worse" attentional states (mean = 3.67%, SD = 2.80%) compared to "better" attentional states (mean = 2.66%, SD = 2.96%; t_{70} = 3.58, p = 0.0006, Cohen's dz = 0.43, 95% CI [0.45, 1.57]). A subsequent analysis focusing specifically on errors for the infrequent food trials revealed that participants also made more errors on these trials during "worse" vs. "better" attentional states (t_{70} = 4.38, p < 0.0001, Cohen's dz = 0.52, 95% CI [0.56, 1.50]).

We therefore replicated prior work in showing that RTs on preceding trials robustly predict the likelihood of an error on the go/no-go task (deBettencourt et al., 2018; Wakeland-Hart et al., 2022). We next asked whether this RT-based measure of attentional states also predicted subsequent recall. There was no difference in recall performance between "better" vs. "worse" attentional states (t_{70} = 1.80, p = 0.077, Cohen's dz = 0.22, 95% CI [-0.13, 2.43]). This

finding is largely in line with our prior work, which showed that faster vs. slower RTs at encoding do not consistently or robustly predict subsequent recall (Jayakumar et al., 2023).

Finally, we examined our main question of interest: differences in temporal organization by attentional state. Based on our prior work (Jayakumar et al., 2023), we expected that RT-based measures of attentional state would not predict the temporal organization of recall. To test this, we conducted a three-way repeated-measures ANOVA on the lag-CRP values with attentional state ("better" vs. "worse" as defined by RT for the preceding three trials), absolute lag (1 to 29), and direction (forward vs. backward) as factors. As expected, and consistent with our previous null findings (Jayakumar et al., 2023), RT-based attentional states did not predict the temporal organization of recall: there was no three-way interaction between attentional state, absolute lag, and direction ($F_{13.69, 958.58} = 1.48$, $p = 0.12$, $\eta_p^2 = 0.02$), nor two-way interactions between attentional state and absolute lag ($F_{15.03, 1052.13} = 1.12$, $p = 0.34$, $\eta_p^2 = 0.016$) or attentional state and direction ($F_{1,70} = 0.16$, $p = 0.70$, $\eta_p^2 = 0.002$). We did, however, observe a significant main effect of absolute lag ($F_{14.22, 995.06} = 9.75$, $p < 0.0001$, $\eta_p^2 = 0.12$), indicating that recall was more likely to transition to items encoded nearby vs. far away, and a main effect of attentional state ($F_{1,70} = 7.89$, $p = 0.0064$, $\eta_p^2 = 0.10$), reflecting superior overall recall for items encoded during better vs. worse attentional states[1].

A follow-up analysis on the ±1 lags, in which forward asymmetry is most pronounced, revealed that recall was more likely to transition forward than backward (+1 CRP > -1 CRP, $F_{1,70}$ = 5.83, $p = 0.02$, $\eta_p^2 = 0.08$); however, this effect did not differ across RT-based attentional

---

[1] This effect is different from the analysis of recall rates across attentional states because recalls that transitioned between an item encoded in a "better" state and an item encoded in a "worse" state were excluded from the lag-CRPs of each attentional state, but not from overall recall for each state, see Methods.

states (no attentional state x lag interaction, $F_{1,70} = 0.25$, $p = 0.62$, $\eta_p^2 = 0.004$; no main effect of attentional state, $F_{1,70} = 1.77$, $p = 0.19$, $\eta_p^2 = 0.025$).

The lag-CRP analysis is therefore consistent with our prior studies in showing that RT-based measures of attention do not predict the temporal structure of memory. To ensure that this null effect was not due to the specific way we used RTs to define attentional states, we repeated the lag-CRP analysis with two other approaches that have been used to define attentional states based on RT.

First, we defined better vs. worse attentional states based on the response time to a particular trial in the encoding task, rather than the preceding three trials. We performed a median split to define "better" attentional states as those with slower RTs and "worse" attentional states as those with faster RTs (Jayakumar et al., 2023). As in our primary response-time-based analyses, and replicating our prior work, we found no effect of RT-based attentional state on recall organization: there was no three-way interaction between attentional state, absolute lag, and direction ($F_{16.62,1163.63} = 1.05$, $p = 0.40$, $\eta_p^2 = 0.015$), nor two-way interactions between attentional state and absolute lag ($F_{15.35,1074.49} = 1.37$, $p = 0.16$, $\eta_p^2 = 0.019$) or attentional state and direction ($F_{1,70} = 0.09$, $p = 0.77$, $\eta_p^2 = 0.001$). Second, we defined better vs. worse attentional states based on RT variability, with trials closer to the mean RT being categorized as a "better" attentional state and trials with RTs that are further away from the mean RT being categorized as a "worse" attentional state (following Esterman et al., 2013; Rosenberg et al., 2013; Jayakumar et al., 2023). Once again replicating our prior work, there was no effect of attentional state on recall organization (ps > 0.13 for all interactions with attentional state).

Together, these results replicate our prior work with RT-based measures of attentional fluctuations(Jayakumar et al., 2023). RT-based measures of attentional state at encoding can

predict online errors and (sometimes) overall recall rates, but they do not predict the temporal organization of subsequent memory (Jayakumar et al., 2023). Conversely, we found that pupil diameter at encoding robustly predicted the temporal organization of memory but not online errors or overall recall. These results suggest that pupil size at encoding, relative to response time measures, may be a more sensitive index of the kinds of cognitive processes that shape the ways memories are organized.

2.3.6 Exploring the relationship between response time and pupil size

Thus far, the results suggest that response time and pupil size may index different components of attention and predict different aspects of cognition. Response time predicts online task performance, whereas pupil size predicts temporal organization of recall. To further investigate the relationship between pupil size and response time measures of attention, we performed an exploratory analysis on the correlation between the two. We expected to find a small, but positive relationship between the two measures of attention.

The inputs to this analysis were the trial-by-trial estimates of pupil size and (preceding trial) RTs that served to define better vs. worse attentional states. We concatenated these residualized and pre-processed pupil sizes and response times across the three blocks for each participant. We then computed the within-participant Pearson correlation coefficient between the two. Next, we Fisher-transformed each participant's Pearson correlation coefficient to ensure normality. Finally, we conducted a one-sample t-test on these Fisher-transformed correlations.

We found a small but statistically significant positive correlation (mean Fisher-transformed r = 0.042) between the response time and pupil size measures (t_{70} = 4.00, p = 0.00016, Cohen's dz = 0.48, 95% CI [0.02, 0.06]). This suggests that response time and pupil size measures of attention capture partly overlapping components of behavior and cognition.

However, the weak relationship shows that there is still a considerable amount of unshared variance between the two measures. This provides further evidence that pupil size and response time may index at least partly different aspects of attention, and in turn can predict dissociable aspects of cognition.

2.4 Discussion

2.4.1 Summary of findings

We examined how spontaneous attentional fluctuations influence the temporal organization of memory. We hypothesized that poor attentional states at encoding will disrupt temporally organized recall, while good attentional states will sustain the temporal organization of memory. We defined attentional states using two complementary measures: pupil size and response times. We found that behavioral and physiological measures of attention predicted different aspects of cognition. The pupil-based measure of attentional states predicted the temporal organization of recall but not online task performance, whereas the response-time-based measure predicted online task performance but not the temporal organization of recall. Together, these results show that the conclusions reached about the effect of attentional fluctuations on memory organization can vary depending on how attention is measured. Fluctuations in pupil diameter are sensitive to cognitive processes that predict the linking of memories through time. On the other hand, response-time-based measures better capture moment-by-moment resource allocation towards an ongoing task. Below, we discuss why these measures of attention may diverge and relate our findings to previous work.

2.4.2 Measuring attention with response times vs. pupil sizes

Response times and pupil size have both frequently been used to index attention (e.g. Kahneman & Beatty, 1966; Richer & Beatty, 1987; Hershman & Henik, 2019), with several

studies finding that these measures are correlated and/or predict similar aspects of cognition and behavior (e.g., Einhauser et al., 2010; de Gee et al., 2014; Unsworth & Robison, 2016; van den Brink et al., 2016; Urai et al., 2017; Unsworth et al., 2018, 2020). However, a small but growing body of work – including our own findings – highlights situations in which pupil size vs. response time measures are dissociable, and predict different components of behavior (e.g., Richer & Beatty, 1987; Hershman & Henik, 2019). Such differences have been observed in a variety of tasks including those involving cognitive control and decision-making (Richer & Beatty, 1987; Hershman & Henik, 2019). One possibility is that pupil diameter and response times index different aspects of processing demands, namely, processing load and processing latency, respectively (Richer & Beatty, 1987; also see Kahneman & Beatty, 1966).

Another possible reason why these measures diverge is that pupil size may more closely track global arousal levels. Pupil size is linked to the activity of the locus coeruleus - norepinephrine (LC-NE) system (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Joshi & Gold, 2020). Higher levels of NE (and high-arousal states) are associated with larger baseline pupil sizes (Einhauser et al., 2010; C.-A. Wang et al., 2018); these states may in turn be linked to memory modulation via the hippocampus (e.g., Harley, 2007; Fastenrath et al., 2014; Madan et al., 2017; Ruivo et al., 2017). Lower levels of NE, associated with low-arousal states and smaller baseline pupil sizes, may promote disengagement from the environment (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010), thus disrupting the ability to link neighboring moments in time.

In addition to the locus coeruleus, pupil size has also been linked to the activity of the basal forebrain cholinergic system (Joshi & Gold, 2020). This link is intriguing because the basal forebrain cholinergic system is associated with attentional processes and memory encoding

(Baxter & Chiba, 1999; Decker & Duncan, 2020), linked to activity in the LC-NE system (Larsen & Waters, 2018), and projects to the hippocampus (Everitt & Robbins, 1997; Semba, 2000) to bias it toward an externally oriented attentional state (Newman et al., 2012; Tarder-Stoll et al., 2020; Poskanzer & Aly, 2023; also see Ruiz et al., 2021). Thus, an interesting possibility to explore in future work is whether variation in pupil size during encoding is linked to fluctuations in cholinergic biasing of hippocampal states toward external attention and memory encoding (Duncan et al., 2012; Duncan & Shohamy, 2016; Aly & Turk-Browne, 2018; Poskanzer & Aly, 2023). If so, further studies could determine whether this cholinergic modulation also predicts the temporal structure of subsequent memory.

Finally, pupil measures and response times may diverge in their predictive power because RTs may be a less pure index of available cognitive resources, as they are inextricably linked to decision difficulty on item judgments (Richer & Beatty, 1987). Pupil measures, particularly the pre-trial baseline measure that is obtained in the absence of visual stimulation or task judgments, may thus be a purer measure of attentional states uncontaminated by decision difficulty or difficulty in perceptual identification. That said, our results are not aligned with the interpretation that one measure of attention (e.g., pupil size) is simply better or more process-pure than another (e.g., response times). This is because both measures predicted aspects of behavior; this suggests that both measures are useful, and they can index complementary components of cognitive function.

2.4.3 Relationship to prior work linking pupil size and memory.

Prior studies have found relationships between baseline pupil measures and overall memory performance (e.g. Hoffing & Seitz, 2014), raising the question of why we did not observe such a relationship. The link between pupil size and memory, however, is not

132

consistently observed. For example, Whitlock et al., 2023 found that pupil size at encoding did *not* predict item recognition memory, but did predict associative memory. Such a result may be consistent with our findings that pupil size does not predict recall of individual items but does predict temporal binding of items encoded close to one another. Future studies can further test the conditions under which pupil size predicts memory by comparing item recognition, associative recognition, and recall in the same experiment.

A recent study using a free recall task examined how pupil dynamics changed during encoding as a function of various study phase manipulations (Unsworth & Miller, 2021). They found that pupil size is sensitive to manipulations of list length and presentation duration of the to-be-encoded items, in a way suggestive of pupil size serving as an index of attentional allocation. This study, however, did not relate these pupil dynamics to subsequent memory or the temporal organization of recall. Moreover, baseline pupil size was used a measure of attention at the level of each list; each list therefore had one pre-list baseline pupil size. Thus, our study builds on this work by obtaining pupil measures of attentional state prior to each individual item, and relating this signal to the temporal structure of recall.

Most relevant to the current work on the temporal organization of memory, prior studies have examined how evoked pupil size at event boundaries is related to subsequent temporal memory for within-event and across-event items (Clewett et al., 2020). The authors employ a dimensionality reduction technique to obtain the various components that make up the evoked pupil response. One of these components - an early peaking signal - predicts subsequent temporal memory. This work complements our own study, which examined pre-trial pupil size rather than evoked pupil size, and temporal contiguity effects in free recall as opposed to memory for relative recency or temporal distance. Together, both the current study and the work

133

of Clewett et al. suggest that pupil measures are powerful in indexing the kinds of cognitive operations that predict the incorporation of temporal context in memory.

In sum, our work joins the body of research linking pupil size to memory by showing a close correspondence between pre-trial pupil size and the ability to link items that were experienced close to one another in time. Our findings therefore go beyond prior studies linking pupil size to overall memory performance (Hoffing & Seitz, 2014) and join a growing body of work illuminating how pupil size may index memory for time and event order (Clewett et al., 2020).

2.4.4 Limitations and future directions

Although we observed our hypothesized effect linking pupil-defined attentional states to memory organization, our study nevertheless has several potential limitations that are worthy of discussion. One such limitation concerns some surprising null effects. For example, prior studies have found relationships between baseline pupil size and online task performance (Unsworth & Robison, 2016; van den Brink et al., 2016) raising the question of why we did not find such a relationship. This difference may be due to the modifications we made to our sustained attention task to make it amenable for testing free recall. Comparison of their tasks vs. ours in a single experiment would allow identification of whether the differences in our results are due to the task or due to differences in eye tracking preprocessing or analysis procedures (such as which trials counted as those being in the poorer attentional state).

Another potential limitation of our study is that recall rates were relatively low, as is typically observed for long study lists (Healey et al., 2019), like the ones used here and in our prior work (Jayakumar et al., 2023). Despite low recall levels, we were able to replicate the signatures of temporally organized recall (Kahana, 1996; Healey et al., 2019). Further, we

showed that attentional states, as defined by pupil size, predict variability in this temporal

structuring despite the low recall rates. Thus, our task and analysis approach were sensitive to

changes in the temporal structure of memory as expressed in free recall. Future studies can

examine the temporal structure of memory using other tasks that allow measurement of

memories that may not be strong enough to be freely generated. For example, further studies can

assess temporal structure in memory by using recency discrimination tasks (Clewett et al., 2020),

cued recall, or temporally structured recognition tests (Schwartz et al., 2005) to allow assessment

of weaker memories.

Our task design differed in several ways from typical studies of sustained attention that

rely on response time measures to index attentional fluctuations. Such studies, for example, may

have shorter stimulus presentations, gradual transitions between images, and repeated (rather

than trial-unique) images. However, despite these differences, we replicated the finding that RTs

from preceding trials predict attentional lapses (deBettencourt et al., 2018; Decker et al., 2022;

Wakeland-Hart et al., 2022); further, the current study and our 4 preceding experiments

(Jayakumar et al., 2023) consistently found that RTs predict errors but not temporal organization

of memory, with varying experimental designs, stimulus presentation conditions, and list lengths.

Together, this suggests that differences between our task and other studies of sustained attention

(e.g., Esterman et al., 2013; Rosenberg et al., 2013; deBettencourt et al., 2018; Decker et al.,

2022; Wakeland-Hart et al., 2022) are unlikely to account for the null relationship between RT

measures and the temporal organization of memory.

Our findings that pupil diameter and response times predict complementary aspects of

behavior raise the question of whether, if combined in the same model, they can do a better job

of predicting diverse types of behavior than either measure alone. Future work can use moment-

by-moment estimates of attentional state "quality" derived from both pupil measures and response times to predict multiple aspects of online behavior and memory.

Further, our finding that moment-by-moment fluctuations in pupil size predict the temporal structure of memory may potentially be linked to findings that memory formation fluctuates rhythmically over time, with superior vs. inferior encoding following a theta rhythm (e.g., Biba et al., 2024). This raises the possibility that the temporal organization of memory encoding may also be rhythmic, with predictable moments of superior vs. inferior temporal binding of experience.

Finally, we focused on pre-trial pupil sizes in the current work, but future studies can use both pre-trial and evoked pupil responses (as examined in Kang et al., 2014; Clewett et al., 2020; Unsworth et al., 2020; Unsworth & Miller, 2021) to determine if they predict different aspects of memory content and organization.

2.4.5 Conclusion

Attentional fluctuations at encoding can shape how memory is later organized. This conclusion, -however, depends critically on how attention is measured. Larger (vs. smaller) pupil sizes at encoding predict superior temporal organization of subsequent memory. Fluctuations in response times predict online task performance but not memory organization. These results show how conclusions about attention and memory may vary substantially depending on whether attention is measured with a behavioral or physiological approach. More broadly, our results join a growing body of work suggesting that behavioral vs. physiological measures of attention can provide complementary insights into cognition.

Chapter 3: Influence of Fluctuations in Attention on Cued Recall

3.1 Introduction

Memories of events are temporally organized: we tend to remember experiences close together in time if they initially occurred close together in time. This temporal organization of memory is influenced by many factors, including list length, presentation time, and age (Healey et al., 2019). We hypothesize that an additional internal factor that shapes the temporal organization of memory is spontaneous attentional fluctuations. These types of fluctuations can occur slowly over the course of the day, but also within shorter periods of time within specific events (Smallwood et al., 2003, 2008; Smallwood & Schooler, 2015). That is, even within a particular task, we regularly fluctuate between periods in which we are focussed on the task at hand and periods in which we "zone out", distracted by our internal thoughts or external events (Smallwood & Schooler, 2006; Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2015)

We addressed the relationship between attentional fluctuations and memory organization in Chapters 1 and Chapter 2 of this dissertation. Participants encoded trial-unique objects and then engaged in a free recall task. We used variability in response times to task questions at encoding, and pre-object baseline pupil size at encoding, to divide trials into those associated with "better" vs. "worse" attentional states. Our main finding was that larger (vs. smaller) pupil sizes at encoding, indicative of "better" attentional states, were associated with stronger temporal organization of subsequent recall. This supported our hypothesis that spontaneous attentional fluctuations, as captured by pupil size, influence the temporal organization of memory.

Our prior studies used a free recall task, which has both strengths and limitations. The primary strength of a free recall task is it allows the measurement of the temporal organization of recall as it naturally unfolds without constraints. This enables precise quantification of several

aspects of memory organization, such as spontaneous temporal clustering in recall and the preferred direction (forward vs. backward) of recall. A main limitation, however, is that free recall primarily tests memory for items that have been encoded well enough to be brought to mind without external aids. We were therefore not able to assess how attentional fluctuations during learning may influence memory for items that were relatively weakly encoded, and unable to be generated freely. Assessment of weaker memories would allow us to determine whether the benefit of good attentional states on memory organization is stronger when more variable memory strengths can be considered, or, alternatively, whether providing a memory aid at retrieval reduces the importance of attention during learning.

We therefore conducted a follow-up study that examines the influence of spontaneous attentional fluctuations on the temporal organization of memory as tested via cued recall. This allows us to assess memories that can be brought to mind with a cue even if they were not able to be freely recalled, and enables us to determine whether the effects of spontaneous attentional fluctuations on the temporal organization of memory generalizes beyond free recall.

Cued recall tests have long been used as a measure of source or associative memory (BÄckman & Nilsson, 1991; Mulligan & Hartman, 1996; Clarke & Butler, 2008; Naveh-Benjamin & Brubaker, 2019). In these studies, participants typically encode items in cue-target pairs. The cue is later presented in the retrieval phase, and participants are asked to recall the target item (BÄckman & Nilsson, 1991; Mulligan & Hartman, 1996; Clarke & Butler, 2008; Naveh-Benjamin & Brubaker, 2019). In the present study, we used a modification of this cued recall task to allow measurement of the temporal organization of memory. Participants encoded a list of object images, as in our previous work. At test, we presented participants with each object image as a cue, and asked them to recall the object that came right after the cue object during

encoding. If they were not able to do so, we instructed them to recall any object that came to mind. In this way, we could test temporal organization of recall by examining the lag and direction between the recalled item and the cued item, based on their relative positions during encoding, while also providing participants with a retrieval cue to aid memory.

By using a cued recall test in this way, we add to the extensive body of work on factors that influence cued recall performance (e.g., Earhard, 1972; Santa et al., 1975; Jones, 1976; Parkin, 1981; BÄckman & Nilsson, 1991; Mulligan & Hartman, 1996; Clarke & Butler, 2008; Criss et al., 2011; Steinmetz et al., 2016; Naveh-Benjamin & Brubaker, 2019). Of most relevance for the current work, the effects of attention on cued recall have primarily been investigated with divided attention manipulations (BÄckman & Nilsson, 1991; Mulligan & Hartman, 1996; Clarke & Butler, 2008; Naveh-Benjamin & Brubaker, 2019). Our current study therefore extends this work by looking at spontaneous attentional fluctuations and by using cued recall tasks to specifically assess the temporal organization of memory.

To do so, we used the same encoding task as we used in Chapter 2, and modified only the recall task. As in Chapter 2, participants encoded trial-unique images of objects while judging each object as food or non-food. Participants' eyes were tracked during encoding, and attentional states were defined using baseline pupil diameter during the 500ms window prior to image onset. Instead of the free recall test we used in prior work (Chapter 1 and 2), we conducted a cued recall test. Each item that was encoded was also used as a cue, with the order of the items at test determined randomly. Participants were instructed to recall the next item in the study phase if they could; if they could not, they were told to recall any item that comes to mind. This approach allows us to examine how attentional fluctuations influence the ability to recall adjacently

encoded items, which were those for which we observed the strongest effects in our prior work (Chapter 2).

We tested two alternative hypotheses regarding the effect of attentional state, defined by pupil size, on cued recall. First, cued recall, by allowing measurement of both strong memories and weaker memories that may not be freely recalled, may allow us to see a large effect of attentional state. That is, memories that are strong enough to be recalled freely may have similar representations even if some were encoded during periods of relatively worse attention. Examining cued recall could therefore allow us to uncover differences that are masked in tests of strong memories. If so, cued recall of the subsequently encoded item should be higher for items encoded with larger vs. smaller pupil sizes, and this effect may be even larger than the one we observed with free recall.

Alternatively, the effects of attentional state on temporal organization of memory may be most pronounced for the strongest memories, which can be freely recalled. Providing a memory cue at retrieval may "rescue" weak memories and reduce the importance of being in a good attentional state at encoding. If so, we may observe a weak or no effect of attentional state on cued recall of upcoming items. This would extend work showing that attentional manipulations at encoding, such as those in divided attention tasks, have stronger effects on free recall than cued recall and recognition memory (for e.g., BÄckman & Nilsson, 1991). Below, we report our preliminary findings, but more data collection is needed before definitively providing evidence for either hypothesis.

3.2 Methods

3.2.1 Participants

We recruited participants using flyers around the Columbia University community and compensated them with cash for their time ($15/hour). To be eligible for the study, participants had to be between 18-35 years of age, have normal or corrected-to-normal vision with contact lenses or LASIK surgery, and provide informed consent in accordance with the Columbia University Institutional Review Board.

Data collection was interrupted by University restrictions on-campus access during the months of April to June. As such, we have N = 15 participants. Participants were aged 18 to 33 years old (Mean = 22.6 years; SD = 3.81), and had 12 - 20 years of education (Mean = 15.8 years; SD = 2.15). Of these participants, 9 identified as female, 4 as male, and 2 as non-binary. In terms of race, 6 participants identified as White, 3 as Black or African American, 5 as Asian, and 1 as multiracial. In terms of ethnicity, 1 participant identified as Hispanic/Latino, and 14 participants as not Hispanic or Latino.

3.2.2 Stimuli

We used the same set of 255 isoluminant images of common objects as in Chapter 2 (see Section 2.2.3.).

3.2.3 Experimental Design and Procedure

The experimental design was similar to the one used in Chapter 2. The study and distraction phases were identical; the only modifications were to the recall phase. Briefly, participants performed the experiment in three blocks, each of each contained three phases: study phase with a go/no-go task, a math distractor phase, and a cued recall phase (Figure 18). Eye-tracking data were acquired during the study phase.
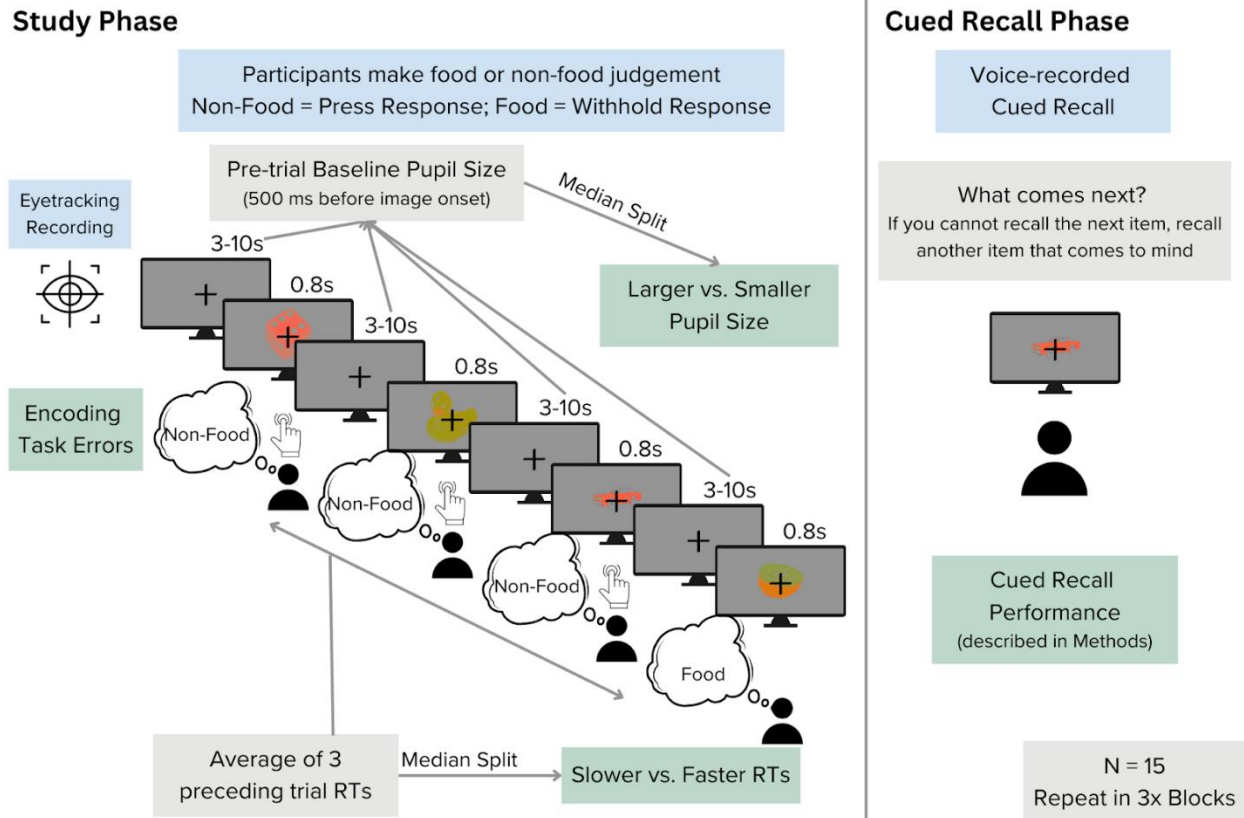
Figure 18. Schematic of the experimental design and task measure. Participants completed the task in three blocks, each of which contained three phases: study phase with concurrent eye-tracking, distractor phase (not pictured), and a cued recall phase. During the study phase, participants were presented with a series of isoluminant images of objects (see Section 3.2.3). For each image, participants were asked to press a button for a non-food item and withhold the button press for a food item. Attentional states were indexed using pre-trial baseline pupil size (median split of the residualized pupil size in the 500ms prior to stimulus onset) and response time (median split of the average of the three preceding trial RTs). Following the math distractor phase (not shown above), participants completed a cued recall phase. Here, participants were cued with an isoluminant image from the corresponding study phase and asked to recall the item that came next from that study phase. If they could not recall the next item, they were asked to recall any other item from the study phase that came to mind. Cued recall performance was computed as described in Methods.

3.2.3.1 Study Phase

During each study phase, participants were presented with 80 trial-unique isoluminant images of objects on the screen, each for 800ms. The inter-trial interval (ITI) was uniformly

distributed between 3 to 10 seconds in increments of 500ms; this ensured that image onsets were unpredictable, thus promoting more fluctuations in attention. A fixation cross was presented in the center of the screen throughout the entire study phase, and participants were asked to fixate on it.

Participants performed a go/no-go task during the study phase - for each presented image, they were asked to judge whether the object depicted in the image was a food or a non-food item. They were asked to press a button if it was a non-food item ("go" trials) and asked to withhold their response if it was a food item ("no-go" trials). The non-food "go" trials occurred 90% of the time while the food "no-go" trials occurred 10% of the time.

3.2.3.2 Distractor Phase

To discourage rehearsal between the study and recall phase, the study phase was followed by a 70-second distractor phase. Participants were presented with 2-digit additions and subtractions and judged whether the solution was correct or incorrect.

3.2.3.3 Cued Recall Phase

Participants performed a cued recall task as the final phase in each of the three blocks. Participants were presented with the image of an object from the study phase as a cue on each trial. They were then asked to recall the object that came next from that study phase. They were told to recall any item that came to mind if they could not recall the next item. For each item, they were given 30 seconds to recall; however they could choose to move forward to the next item after 5 seconds. Recall was voice-recorded with a microphone, and later transcribed and scored.

3.2.3.4 Practice Block

Participants performed a practice block with the three phases (study, distractor, cued recall) before the first experimental block. The practice study phase contained 15 items, 4 of which were food items. The practice math and recall phases were identical to the experimental blocks. The data from the practice block were not analyzed further.

3.2.4 Pupil Data Acquisition

Pupil measures were acquired using an EyeLink 1000 Plus eye tracker by SR Research. Participants used a head-stabilized mount while their right eye was tracked at 1000Hz. They were seated 93cm away from the computer where the experimental task was presented. A 24-inch LED monitor with resolution of 1920px x 1080px and a refresh rate of 60Hz was used to display the experimental task and stimuli. Participants performed a 9-point calibration and validation procedure before beginning the experiment. They were explicitly instructed to remain as still as possible and fixate on the cross that appeared in the center of the screen throughout the study phase.

3.2.5 Analyses

3.2.5.1 Eye tracking data pre-processing

Eye tracking measures were processed using the same pipeline as Chapter 2 (see Section 2.2.6.1). Briefly, we converted the raw data from EDF to .asc format, and used 'eyeblinker' R package to read the .asc files. We then extended eye blinks by 100ms on either side, performed pupil foreshortening effect correction, interpolated for blinks and off-screen fixations, and obtained residuals by regressing out trial numbers from pupil size within participants. These residualized pupil data were used for further analyses.

3.2.5.2 Pupil measures as an index of attention

We used the residualized pupil size from the pre-processing pipeline described above as an index of attentional state. We averaged the residual pupil size in the 500ms window before image onset and then performed a median split of these mean pupil sizes to create two states: larger vs. smaller pupil sizes (van den Brink et al., 2016; Unsworth et al., 2020). Larger pupil sizes were taken as an index of better attentional states and. smaller pupil sizes an index of worse attentional states.

3.2.5.3 Errors at encoding

We compared the errors participants made in the encoding task during the two attentional states. We calculated errors as the percentage of incorrect responses for the food vs. non-food judgment. Incorrect responses included incorrect button presses to the "no-go" (food) trials and incorrect withholdings to the "go" (non-food) trials. We then conducted a paired samples t-test to examine whether errors differed between the better vs. worse attentional states as indexed by larger vs. smaller pupil size at encoding.

3.2.5.4 Cued Recall Performance

During the cued recall test, participants were asked to recall the next item from the study phase; if they could not recall the next item, they were asked to recall any item that came to mind. This allowed us to examine temporal organization at recall by calculating the distance between the cue item and the recalled item in terms of their positions on the study list. Recalled items that were encoded after the cue item in the study list were classified as a forward direction of recall (positive lags); recalled items that were encoded before the cued item were classified as a backward direction of recall (negative lags). The numeric value of the lag was calculated as the difference in study-list serial positions for the cue item and the recalled item. Thus, a correctly

recalled next item would be at a distance (or lag) of +1 from the cued item based on their study

positions. For each participant, we calculated the number of items they correctly recalled at each

lag. We then divided each of these values by the total number of study-list items they recalled to

yield the percentage of all recalled items from the study list that fell at each lag.

Our primary hypothesis was that recall performance for the next item would be higher for

the attentional state indexed by larger vs. smaller pupil sizes. To further explore the data,

however, we also examined recall differences at lags of -1, +2, -2, and, additionally collapsed

across lags 1 to 5 and lags -1 to -5.

To examine differences between the "better" (larger pupil size) and "worse" (smaller

pupil size) attentional states, we classified each recall trial as belonging to a "better" state,

"worse" state, or a "transition" based on attentional states at encoding. If both the cue item and

the recalled item were associated with larger pupil sizes at encoding, the trial was considered to

fall in the "better" state. If both items were associated with smaller pupil sizes, the trial was

classified as the "worse" state. If the items belonged to different attentional states at encoding,

the trial was considered a "transition" trial and was not analyzed further. We compared recall

across states for both: (1) overall recall (total number of study-list items recalled from each state

divided by the total number of items encoded in that state); and (2) recall as a function of lag

(total number of study-list items recalled at that lag for that state, divided by the total number of

study-list items recalled from that state). We conducted group-level analyses comparing the

"better" and "worse" states with a paired-samples t-test.

3.3 Results

As noted earlier, data collection was interrupted by restrictions on campus access in Spring 2024. Below, we report preliminary analyses on our dataset so far; further data collection will be essential for a well-powered study.

3.3.1 Online task performance

We first examined whether fluctuations in attentional states were associated with errors in the go/no-go task during encoding. Performance on the food or non-food judgment task was very high; mean error was 4.78% (SD = 3.27; median = 4.17%). There was no significant difference in errors between the two attentional states, as defined by larger (mean = 5.27%, median = 3.60%, SD = 4.52) vs. smaller (mean = 4.43%, median = 3.60%, SD = 3.44) pupil sizes (t_{14} = 0.71, p = 0.49, Cohen's dz = 0.18, 95% CI [-1.68,3.36]). Thus, pupil size at encoding did not predict online task performance in our current sample.
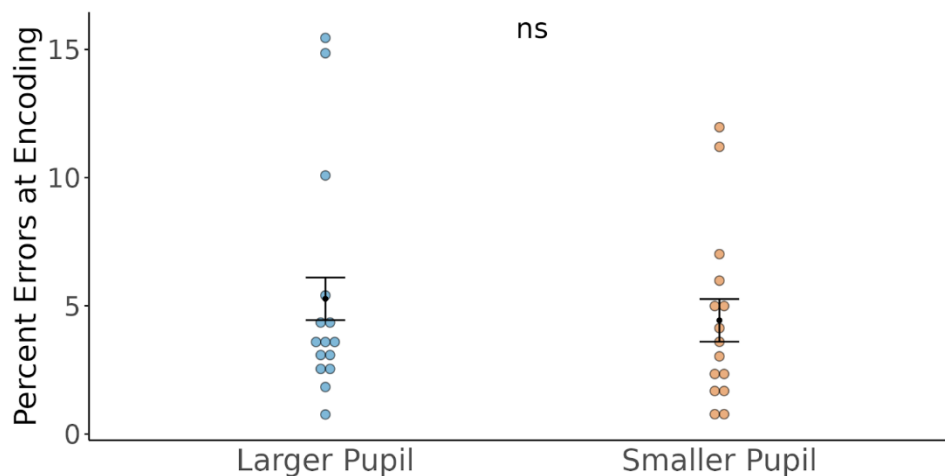
Figure 19. Encoding task performance by pupil-based attentional states. Individual points show the percentage of encoding judgment errors made by each participant during the "better" (larger pupil) and "worse" (smaller pupil) attentional states. Block points indicate the mean of the encoding errors; error bars indicate the standard error of the within-participant difference between the "better" and "worse" attentional states. There was no significant difference in the online task performance between the two attentional states. ns = not statistically significant.

3.3.2 Cued recall performance

Overall, participants correctly recalled an item from the study list 59.89% of the time (SD = 20.41%, median = 64.58%). Of these, participants correctly recalled the next item 13.85% of the time (SD = 16.01%, median = 6.60%).

Our primary hypothesis was that cued recall for the next item from the encoding list would be superior when both items were encoded in a better vs. worse attentional state.

Before testing our main hypothesis, we first examined overall recall performance as a function of attentional state at encoding – regardless of the lag between the cue item and the recalled item. Overall recall performance did not differ by attentional state at encoding (larger pupil size: Mean recall = 30.67%, SD = 13.83%, median recall = 31.67%; smaller pupil size: Mean recall = 32.39%, SD = 13.65%, median = 33.33%; $t_{14} = 0.46$, $p = 0.65$, Cohen's $dz = 0.12$, 95% CI [-6.24, 9.68]).

We next visualized the temporal organization of cued recall by plotting a curve depicting recall performance as a function of the lag (or distance) between the cue item and recalled item at encoding. Such a curve is similar to the lag-CRP curves from Chapter 2 (see Section 2.2.6.6.). We obtained the percentage of times participants recalled an item at a given lag from the cued item, separately for item pairs that were both encoded during a "better" (larger pupil size) vs. "worse" (smaller pupil size) attentional state at encoding. Because participants were instructed to recall the next item, we expected a peak in this curve at lag +1. Indeed, recall curves for both attentional states resembled the typical shape of a free recall lag-CRP curve (Figure 20). There was a peak in cued recall at lag +1, as expected, indicating that participants were successful in following instructions to recall the next item. Further, the close resemblance of our cued recall

curves to the lag-CRP curves obtained via free recall (see Chapters 1 and 2) indicates that our modified cued recall procedure was effective at capturing the temporal structure of memory.
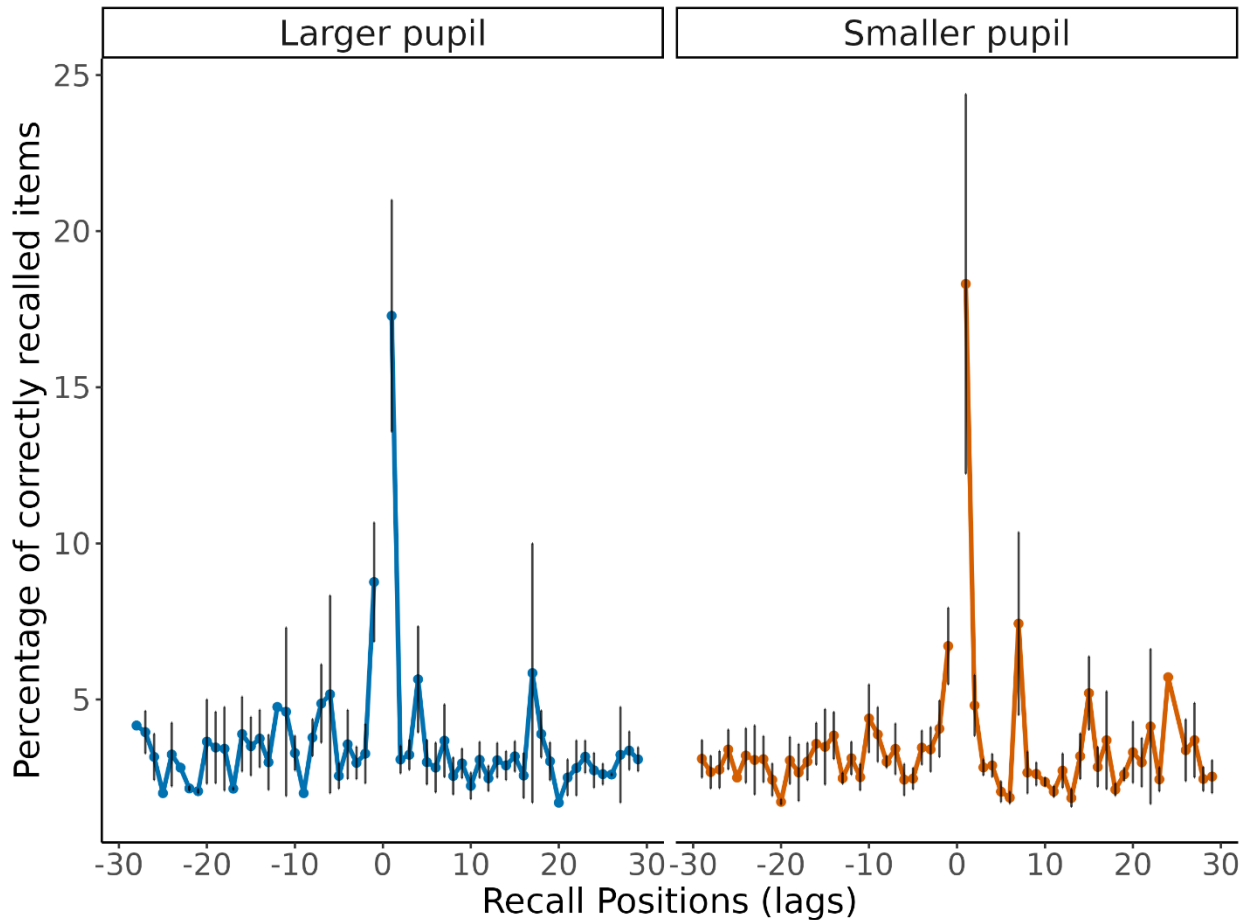
Figure 20. Lag-Percentage of Correctly Recalled Items curves by pupil-based attentional state. We visualize the temporal organization of cued recall by plotting a curve with lag on the x-axis and the percentage of correctly recalled items on the y-axis (see Results). These curves are plotted separately for items encoded in the better (larger pupil; left) and worse (smaller pupil; right) attentional states. As expected, there is a peak in the cued recall at lag +1. Furthermore, this curve resembles the typical shape of free-recall lag-CRP curves from Chapters 1 and 2. Error bars represent the standard error.
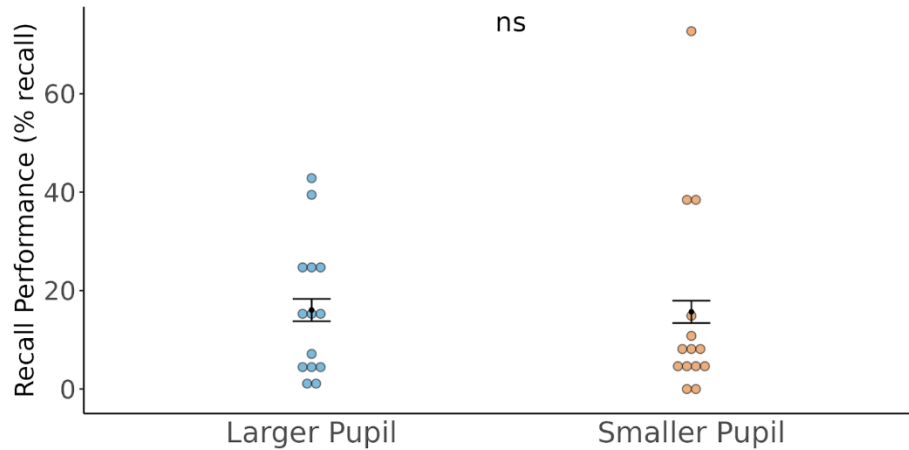
Figure 21. Recall performance at lag +1 by pupil-based attentional states. Individual points show the recall performance at lag +1 (i.e. next item performance) for each participant as a function of whether they were encoded during better (larger pupil) or worse (smaller pupil) attentional states. Black points indicate the mean of the encoding errors; error bars indicate the standard error of the within-participant difference between the "better" and "worse" attentional states. There was no significant difference in cued recall performance of the next item between the two attentional states. ns = not statistically significant.

Having confirmed that our task design was effective, we next turned to comparing cued recall across attentional states. We hypothesized that recall performance at lag +1 would be higher for the "better" attentional state vs. the "worse" attentional state. However, we found no significant difference between states at lag + 1 (larger pupil size: mean recall = 17.30%, SD = 13.37%, median = 16%; smaller pupil size: mean recall = 18.31%; SD = 21.05%, median = 9.07%; t_{13} = 0.11, p = 0.9, Cohen's dz = 0.03, 95% CI [-6.55, 7.27]).

We further explored recall performance at lags of +2, -1, and -2. However, we did not find any significant differences in recall performance between the two attentional states at these lags (lag +2: t_{11} = 1.34, p = 0.21, Cohen's dz = 0.39, 95% C.I. [-0.99, 4.06]; lag -1: t_{12} = 1.18, p = 0.26, Cohen's dz = 0.33, 95% C.I. [-1.77, 5.94]; lag -2: t_7 = 0.21, p = 0.84, Cohen's dz = 0.074, 95%CI [-3.19, 3.81]). Similarly, we also tested for differences in recall performance collapsed across the first five forward lags (+1 to +5) and first five backward lags (-1 to -5). Here too, we did not find any differences in recall performance between the two states (forward: t_{14} = 1.10, p =

150

0.29, Cohen's dz = 0.28, 95% CI [-2.30,7.15]; backward: t_{14} = 0.81, p = 0.43, Cohen's dz = 0.21, 95% CI[-1.79, 6.55]).

Together, these preliminary results show that our modified cued recall task was effective in detecting temporal organization in memory. However, our results thus far suggest that attentional states at encoding do not influence the temporal organization of cued recall. Nevertheless, we must be cautious in interpreting these results given the low number of participants.

3.4 Discussion

We examined how endogenous fluctuations in attention influence the temporal organization of memory when cues are provided at recall. We defined attentional states at encoding using pupil size. We tested two competing hypotheses. First, given that cued recall may allow us to examine both strongly and weakly encoded memories, whereas free recall may primarily tap stronger memories, we may observe a larger effect of attentional fluctuations on cued recall, compared to our prior work with free recall (see Chapter 2). That is, memories that are strong enough to freely recall may not show robust differences in their representations based on attentional states at encoding; examining cued recall could allow us to uncover differences that are masked in tests of strong memories. Alternatively, the cue at recall may help rescue weakly encoded memories or those encoded in a worse attentional state, thus resulting in a small or null effect of attentional states on cued recall.

In preliminary analyses, we found that pupil size at encoding did not predict online task performance or cued recall performance. However, given the pause in data collection due to protest-related shut-downs at Columbia University and the resulting low number of participants (N = 15), these results should be interpreted with caution. At this time, there is insufficient

evidence to provide strong support for either hypothesis. Given the preliminary nature of our results, below we outline our plans and potential conclusions upon completing data collection.

If we were to find that pupil size at encoding was associated with differences in cued recall performance, we would directly compare the effect sizes for the influence of attentional state on cued recall with the effect size for free recall (from Chapter 2). If we find that the cued recall effect is larger than the one for free recall, it would provide support for the hypothesis that the effects of attention are more robust when weaker memories can be measured. However, if we find that the cued recall effect is smaller than the effect for free recall, or if we find no effect of attention on cued recall performance at all, it would provide support for the alternative hypothesis. Such a result would demonstrate that providing a cue at recall may rescue memories that were encoded weakly in the worse attentional state, thus dampening the observed effects of attentional fluctuations on the temporal organization of memory.

If we find an effect of attentional states on cued recall, it would also indicate that our results from Chapter 2 are generalizable to other memory tests. Alternatively, if we do not find an effect, the null findings may be partly attributable to more interference during cued recall vs. free recall. In the cued recall test, participants see every image from the study list again at recall, and each item serves both as a cue and an item to be recalled. Because items are presented in a random order at the test, the items preceding and succeeding a given item are likely different from those at encoding, potentially leading to interference in memory for the surrounding items. This possibility would be in line with research on retroactive interference, which shows that learning new information interferes with already encoded information (Unsworth et al., 2013). One way to circumvent this limitation in future work may be to randomly choose a subset of objects from the encoding list such that a given item is either a cue or a target but not both.

Another possibility is to present items in the test in the same order as they were presented at study; however, this reinstatement of temporal context at the test may provide a strong aid to memory – which may in turn overpower the subtle effects of attentional state at encoding.

In sum, we used a cued recall test to assess how the effects of attentional fluctuations on the temporal organization of memory may vary when weaker memories are probed. The results of this study will allow us to adjudicate between two competing hypotheses about how memory cues may enhance or blunt the effects of attention on temporally organized recall. Completion of the study will allow us to assess how cueing interacts with attention to influence the temporal organization of memory.

Discussion

Dissertation Overview

In this dissertation, I examined how spontaneous fluctuations in attention influence the temporal organization of memory. I created an experimental design incorporating prior research from two areas typically studied separately - sustained attention and episodic memory. Across the three chapters, I examined the hypothesis that attentional fluctuations shape the temporal organization of memory such that optimal attentional states support the temporal structure of memories, while sub-optimal states hamper them.

In Chapter 1, I introduced the experimental design in a series of four behavioral studies, replicating prior research demonstrating that response-time-based measures of attentional states predict online task performance. However, I did not find a relationship between these response-time-based measures of attentional fluctuations and the temporal organization of memory. The results suggested that either the temporal organization of memory was robust to attentional fluctuations, or that response time did not capture the kinds of attentional fluctuations that influence the temporal organization of memory. I tested these hypotheses in Chapter 2.

In Chapter 2, I modified the experimental design and tested the alternate hypotheses by characterizing attentional fluctuations using two complementary measures: pupil size and response time. I found that pupil-based measures of attentional states predicted the temporal structure at recall, but not online task performance; response-time-based measures, on the other hand, predicted only the online task performance. These results provide two important conclusions. First, this is the first empirical evidence demonstrating that attentional fluctuations, as indexed by pupil size, impact the temporal structure of memory through its influence of temporal contiguity and forward asymmetry effects. Furthermore, by showing that behavioral vs.

physiological measures of attention predict different aspects of cognition and may be indexing distinct components of attention, these results underscore the importance of using complementary measures of cognition.

I extend these findings in Chapter 3 by examining how retrieval cues interact with the attentional state at encoding to influence the temporal organization of memory. Preliminary results suggest that attentional fluctuations do not impact the temporal organization of memory when a cue is presented at retrieval. However, it is too early to interpret this data with the low number of participants due to Columbia University campus closures from April to June and the resulting research pause. Completing this study will provide evidence for the role of retrieval cues in boosting or weakening the influence of attentional fluctuations of memory.

By combining typically disparate areas of research, this dissertation provides foundational evidence for the influence of attentional fluctuations on the temporal organization of memory. The results suggest that suboptimal attentional states disrupt the temporal organization of memory, while optimal states sustain this organization. Further, it adds to the growing body of research (Kahneman & Beatty, 1966; Hershman & Henik, 2019) showing the dissociation between behavioral and physiological measures in predicting human behavior and encourages the use of complementary measures to capture internal cognitive states. Together, this work expands our understanding of the various factors influencing memory organization and identifies endogenous fluctuations in attention as one of them.

Disentangling Arousal and Attention

Pupil size has been used extensively in cognition research as an indicator of both arousal and attention (Aston-Jones & Cohen, 2005; Laeng et al., 2012; Mathot, 2018; Joshi & Gold, 2020; Strauch et al., 2022). One proposal theorizes that pupil size may be a marker for the

activity of the locus coeruleus norepinephrine (LC-NE) system, thus providing insight into global arousal levels (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Mathot, 2018; Joshi & Gold, 2020). While the exact mechanism remains to be elucidated, activity in the locus coeruleus during high arousal states is associated with higher levels of norepinephrine and observed dilation of the pupil (Gilzenrat et al., 2010; Larsen & Waters, 2018). Furthermore, LC-NE system has direct projections to the dentate gyrus in the hippocampus and LC activity has been associated with learning and memory (Haring & Davis, 1985; Patton & McNaughton, 1995; Samuels & Szabadi, 2008; Kempadoo et al., 2016; Wagatsuma et al., 2018; Kaufman et al., 2020), thus implicating the role of arousal-based pupil signals in mnemonic processes.

Pupil dilation may also be modulated by the basal-forebrain acetylcholine system, either independently or in concert with the LC-NE system (Larsen & Waters, 2018; Joshi & Gold, 2020). The basal-forebrain acetylcholine system also projects to the hippocampus (Everitt & Robbins, 1997; Semba, 2000) with the acetylcholine levels in the hippocampus biasing internal vs. external attention (Newman et al., 2012; Poskanzer & Aly, 2023; see also Tarder-Stoll et al., 2020). Thus, pupil-based signals may also be linked to the role of this basal-forebrain acetylcholine system in memory processes.

Alternatively, pupil diameter may reflect attention-related processes (Strauch et al., 2022). This view is supported by early research capturing the modulation of pupil measures through manipulations of mental effort, processing demands, and capacity of attention (Hess & Polt, 1960, 1964; Kahneman & Beatty, 1966; Kahneman, 1973; Beatty, 1982; see Laeng et al., 2012 for a review of these studies). Strauch et al. (2022) propose a novel model that relates several low, intermediate, and high-level factors to pupil measures. While the low-level factors deal with visual properties (luminance and viewing distance) that we typically control for in

pupillometry studies, the intermediate-level (alerting and orienting) and high-level factors (executive functioning) overlap with the framework of attention networks proposed by Petersen & Posner (2012) and are relevant to our discussion. The alerting system, associated with alertness and vigilance, maps onto the LC-NE pathway similar to arousal (Petersen & Posner, 2012; Strauch et al., 2022). Furthermore, the executive function system, associated with the allocation of focal attention and activity in the frontoparietal networks, is related to the efficient processing of sensory information (Strauch et al., 2022). Ultimately, both these factors may be involved in pupil-related changes during encoding.

Altogether, pupil measures are likely a mix of arousal and attention. Indeed arousal and attention are inextricably linked (Petersen & Posner, 2012; Strauch et al., 2022), particularly when examining fluctuations in sustained attention (e.g. Parasuraman, 2000; Foucher et al., 2004; Oken et al., 2006). Whether pupil size may be interpreted as an arousal-related or an attention-related signal may depend on the type of task used and the control of external factors (Mathot, 2018; Strauch et al., 2022). Future research could disentangle the many factors contributing to arousal- vs. attention-related pupil signals and elucidate the overlapping and distinct neural mechanisms underlying these pupil changes.

Attention and Memory Performance

The relationship between attention and memory has been extensively studied in recent decades, showing a robust relationship between attention at encoding and subsequent memory (see Chun & Turk-Browne, 2007; Aly & Turk-Browne, 2017). Prior research on divided attention has shown a relationship between dual-tasking and worse subsequent memory (e.g. Craik et al., 1996; Anderson et al., 1998; Craik et al., 2018). In addition, RT-based attentional states are seen to predict subsequent recognition memory performance (deBettencourt et al.,

2018; Wakeland-Hart et al., 2022) as do pupil-based measures of attention (e.g. Hoffing & Seitz, 2014). Based on these findings, one might predict that attentional fluctuations during encoding, whether indexed by RT or pupil measures, will predict recall performance. However, across the three chapters in this dissertation, we found that response-time-based attentional states predicted recall performance only in Experiment 4 of Chapter 1 (see Section 1.5). Even pupil size which predicted the temporal organization of recall did not predict recall performance (see Chapter 2).

To get at this surprising result, first, lets compare the tasks in this dissertation to the ones from the divided attention studies. Dual-tasking at encoding involves performing two tasks simultaneously where both require complete attention and focus. This is seen as a decrement in task performance on both tasks at encoding. In contrast, attentional fluctuations at encoding may be more subtle and potentially less disruptive to the primary task. Even though there is a decrease in online task performance during these worse attentional states, participants are still performing reasonably well on the encoding task. Future research could compare the decrement in online task performance for conditions with divided attention vs. attentional fluctuations to test this possibility.

Why do we see no differences in memory performance when other studies examining sustained attention do? This effect of attentional states on memory does not seem to be consistent. For example, a recent study showed that pupil dilation predicts subsequent memory in an associative memory task, but not in an item recognition task (Whitlock et al., 2023). This is in line with our findings from this dissertation. It may be the case that pupil size does not predict item memory, but may predict the binding of items encoded close together, as in both recall and associative memory. This is an open question that can be addressed in future research.

A final possibility is to examine if task-evoked pupil responses at encoding may be more indicative of recall performance. Thus far, this dissertation has only examined baseline pupil measures. Many studies have shown an increase in evoked response associated with subsequent item recognition (e.g., Kafkas & Montaldi, 2011; Papesh et al., 2012; Naber et al., 2013; Ariel & Castel, 2014), even in associative memory tasks (Hoffing & Seitz, 2014). Further, factor analysis of the evoked pupil response has shown an early peaking component associated with temporal memory (Clewett et al., 2020). Thus, it is plausible that evoked pupil measures may be tracking some type of encoding strength, which would be in line with the phasic mode of the LC-NE system functioning (Aston-Jones & Cohen, 2005; Joshi & Gold, 2020). Future research in this area can disentangle the contribution of baseline vs. evoked pupil measures to subsequent memory retrieval and organization.

Attention and Memory Organization

The primary finding from this dissertation is that fluctuations in attention at encoding, as indexed by baseline pupil measures, influence subsequent temporal organization at retrieval. These findings are in line with prior work showing that pupil measures track event boundaries; these boundary-related pupil dilations predict later temporal order memory (Clewett et al., 2020). An exploratory analysis showed that the variability of pupil size between two items at encoding was related to subsequent temporal recency judgments (Clewett et al., 2020). This was a key finding demonstrating that pupil measures at encoding may be sensitive to incorporation of temporal information in memory. Chapter 2 and ongoing work from Chapter 3 build on these prior findings. First, by using baseline measures of pupil size, we were able to obtain a measure of the internal cognitive state independently from the trial response or the onset of visual stimulus. Second, using lag-CRP curves from free recall, allowed us to quantify not only the

temporal organization of recall. This provided insight into the temporal dynamics of recall by quantifying temporal contiguity and forward asymmetry effects. Thus, my dissertation, and particularly Chapter 2, provides the first empirical evidence of pupil-linked signals tracking mnemonic temporal organization through the influence of temporal contiguity and forward asymmetry effects.

A recent framework examining event segmentation (the process of organizing our continuous experiences into events) has proposed that changes in internal cognitive states may also create event boundaries (Y. C. Wang et al., 2023). Prior research has shown that externally triggered event boundaries shape temporal memory such that items within an event are more temporally structured than items across events (Ezzyat & Davachi, 2010; DuBrow & Davachi, 2013, 2014; Ezzyat & Davachi, 2014; DuBrow & Davachi, 2016; Heusser et al., 2018). If internal changes in cognitive states also act as event boundaries, then we would see better temporal memory within the same internal state than across distinct internal states. I explored this in Chapter 1 (section 1.2.1.4.5, Fig 2) by examining how likely participants were to transition between items within the same segment of the attentional state, between two adjoining segments of different attentional states, or between two different segments of the same attentional state. Across the four studies in Chapter 1, we did not find any differences in the transition types by response-time-based attentional states. However, this is an interesting question that could be explored in future research with pupil-based attentional states.

Finally, this dissertation focused on the temporal organization of memory as the first step toward connecting fluctuations in attention with episodic memory organization. However, there are other types of mnemonic organization, including semantic organization (Eichenbaum, 2017). Indeed, there is recent evidence that pupil measures at encoding are related to subsequent source

or associative memory retrieval (Whitlock et al., 2023). An area for future research could be an exploration of how these pupil-linked attentional processes at encoding influence the semantic organization at retrieval. An additional question may be how semantic and temporal organization interact at retrieval, particularly for events encoded during worse attentional stats.

Multiple measures of human behavior

Many studies use both response times and pupil measures to index attention (e.g., Kahneman & Beatty, 1966; Richer & Beatty, 1987; Hershman & Henik, 2019) and predict other measures of behavior (e.g., Einhauser et al., 2010; de Gee et al., 2014; van den Brink et al., 2016; Unsworth & Robison, 2016; Urai et al., 2017; Unsworth et al., 2018, 2020). Among these, a small number of them have also highlighted the importance of using complementary measures to study behavior given the observed dissociations between response times and pupil measures in predicting human behavior (e.g. Richer & Beatty, 1987; Hershman & Henik, 2019). Richer & Beatty (1987) suggest that this dissociation may be due to the different aspects of processing demands being indexed by the two measures; pupil diameter may be providing insight into processing load while response times may be probing processing latency (also see Kahneman & Beatty, 1966).

Another reason for this dissociation may be specific tasks and what response it elicits from the participant (Hershman & Henik, 2019). In the studies in this dissertation, participants responded with a button upon making a judgment decision on each image. Thus, response times for the button press include an underlying decision-making process (Richer & Beatty, 1987), along with the task's cognitive load or processing demands. On the contrary, the pupil measures - particularly, the pre-image baseline measures used in this dissertation - may be a more accurate measure of the attentional state leading up to stimulus onset, without involving the decision-

161

making component. Nevertheless, both measures predicted different aspects of behavior, thus highlighting the importance of indexing cognitive processes using multiple measures.

This use of multiple measures could also extend to include neural data. For instance, some studies have used a combination of EEG, fMRI, and pupillometry to predict subsequent response times (Hong et al., 2023; Makovac et al., 2019), activity in the fronto-parietal network (Hong et al., 2023), and mind-wandering on thought probes (Makovac et al., 2019). Building on these studies, future research could examine pupil-dilation-related activity in the locus coeruleus or the basal forebrain to understand which region may be a better predictor of response times, online task performance, or subsequent memory organization. This may allow us to adjudicate between the neuromodulatory mechanisms underlying arousal-related pupil changes (as discussed in the section 'arousal vs. attention').

Similarly, a future direction may include examining hippocampal representational patterns as measures of attentional states. Recent work has shown that during relational processing, hippocampal patterns are stable within distinct attentional states (Aly & Turk-Browne, 2016b). Future work could examine whether the stability of the hippocampal activity patterns is associated with better vs. worse attentional states as defined by pupil size and whether these patterns, either independently or together with pupil measures, predict memory organization.

Conclusion

In this dissertation, I examined how spontaneous fluctuations in attention shaped the temporal organization of memory. Across three chapters, I showed that temporal organization at retrieval is shaped by fluctuations in pupil-based attentional states at encoding such that worse attentional states disrupt the temporal structure of memory, while better attentional states sustain

the temporal structure. I also demonstrated that behavioral and physiological measures of attention provide complementary insights into cognition. Together, these findings advance our understanding of the complex interactions between attentional states and subsequent memory and contribute to our knowledge about factors that influence the temporal organization of memory.

References

Aly, M., & Turk-Browne, N. B. (2016a). Attention promotes episodic encoding by stabilizing
hippocampal representations. *Proceedings of the National Academy of Sciences*, *113*(4),
E420–E429. https://doi.org/10.1073/pnas.1518931113

Aly, M., & Turk-Browne, N. B. (2016b). Attention Stabilizes Representations in the Human
Hippocampus. *Cerebral Cortex*, *26*(2), 783–796. https://doi.org/10.1093/cercor/bhv041

Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by,
attention. In *The Hippocampus from Cells to Systems: Structure, Connectivity, and
Functional Contributions to Memory and Flexible Cognition* (pp. 1–589).
https://doi.org/10.1007/978-3-319-50406-3

Aly, M., & Turk-Browne, N. B. (2018). Flexible weighting of diverse inputs makes hippocampal
function malleable. *Neuroscience Letters*, *680*, 13–22.
https://doi.org/10.1016/j.neulet.2017.05.063

Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). *The Attentional Demands of
Encoding and Retrieval in Younger and Older Adults: 1. Evidence From Divided
Attention Costs*.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla
in our midst: An online behavioral experiment builder. *Behavior Research Methods*,
52(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively
studying important information. *Experimental Brain Research*, *232*(1), 337–344.
https://doi.org/10.1007/s00221-013-3744-5

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS

    COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal

    Performance. *Annual Review of Neuroscience*, *28*(1), 403–450.

    https://doi.org/10.1146/annurev.neuro.28.061604.135709

BÄckman, L., & Nilsson, L.-Gö. (1991). Effects of divided attention on free and cued recall of

    verbal events and action events. *Bulletin of the Psychonomic Society*, *29*(1), 51–54.

    https://doi.org/10.3758/BF03334767

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from

    long-term memory. *Journal of Experimental Psychology: General*, *113*(4), 518–540.

    https://doi.org/10.1037/0096-3445.113.4.518

Bainbridge, W. A. (2019). Chapter One - Memorability: How what we see influences what we

    remember. In K. D. Federmeier & D. M. Beck (Eds.), *Psychology of Learning and*

    Motivation (Vol. 70, pp. 1–27). Academic Press.

    https://doi.org/10.1016/bs.plm.2019.02.001

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory

    separate from attention and priming. *Neuropsychologia*, *141*, 107408.

    https://doi.org/10.1016/j.neuropsychologia.2020.107408

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs.

    Journal of Experimental Psychology: General, *142*, 1323–1334.

    https://doi.org/10.1037/a0033872

Bainbridge, W. A., & Oliva, A. (2015a). A toolbox and sample object perception data for

    equalization of natural images. *Data in Brief*, *5*, 846–851.

    https://doi.org/10.1016/j.dib.2015.10.030

Bainbridge, W. A., & Oliva, A. (2015b). Interaction envelope: Local spatial representations of

objects at all scales in scene-selective regions. *Neuroimage*, *122*, 408–416.

https://doi.org/10.1016/j.neuroimage.2015.07.066

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews,

T. J. (2020). Power contours: Optimising sample size and precision in experimental

psychology and human neuroscience. *Psychological Methods*, *26*(3), 295.

https://doi.org/10.1037/met0000337

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called

Psychological Science. *Perspectives on Psychological Science*, *7*(6), 543–554.

https://doi.org/10.1177/1745691612459060

Baxter, M. G., & Chiba, A. A. (1999). Cognitive functions of the basal forebrain. *Current

Opinion in Neurobiology*, *9*(2), 178–183. https://doi.org/10.1016/S0959-4388(99)80024-

5

Beatty, J. (1982). *Task-Evoked Pupillary Responses, Processing Load, and the Structure of

Processing Resources*.

Biba, T. M., Decker, A. L., Herrmann, B., Fukuda, K., Katz, C., Valiante, T., & Duncan, K.

(2024). *Memory's pulse: Episodic memory formation is theta-rhythmic*. OSF.

https://doi.org/10.31234/osf.io/s8nda

Brand, A., Bradley, M. T., Best, L. A., & Stoica, G. (2008). Accuracy of effect size estimates

from published psychological research. *Perceptual and Motor Skills*, *106*(2), 645–649.

https://doi.org/10.2466/pms.106.2.645-649

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase

II: 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953.

https://doi.org/10.1371/journal.pone.0106953

Burlingham, C. S., Mirbagheri, S., & Heeger, D. J. (2022). A unified model of the task-evoked

pupil response. *Science Advances*, *8*(16), eabi9979.

https://doi.org/10.1126/sciadv.abi9979

Chan, S. C. Y., Applegate, M. C., Morton, N. W., Polyn, S. M., & Norman, K. A. (2017).

Lingering representations of stimuli influence recall organization. *Neuropsychologia*, *97*,

72–82. https://doi.org/10.1016/j.neuropsychologia.2017.01.029

Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious

awareness and everyday cognitive failures. *Consciousness and Cognition*, *15*(3), 578–

592. https://doi.org/10.1016/j.concog.2005.11.009

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A

bidirectional state model of task engagement/disengagement and attention-related errors.

Cognition, *111*(1), 98–113. https://doi.org/10.1016/j.cognition.2008.12.009

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-

wandering as spontaneous thought: A dynamic framework. *Nature Reviews

Neuroscience*, *17*(11), 718–731. https://doi.org/10.1038/nrn.2016.113

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2010). A Taxonomy of External and

Internal Attention. *Annual Review of Psychology*, *62*(1), 73–101.

https://doi.org/10.1146/annurev.psych.093008.100427

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Clark, D. P., & Bruno, D. (2021). Time is of the essence: Exploring temporal and spatial organisation in episodic memory. *Quarterly Journal of Experimental Psychology*, *74*(8), 1406–1417. https://doi.org/10.1177/1747021821993823

Clarke, A. J. B., & Butler, L. T. (2008). Dissociating word stem completion and cued recall as a function of divided attention at retrieval. *Memory*, *16*(7), 763–772. https://doi.org/10.1080/09658210802261116

Clewett, D., & Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences*, *17*, 186–193. https://doi.org/10.1016/j.cobeha.2017.08.013

Clewett, D., Gasser, C., & Lila, D. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(1). http://dx.doi.org/10.1038/s41467-020-17851-9

Cortis, C., Dent, K., Kennett, S., & Ward, G. (2015). First things first: Similar list length and output order effects for verbal and nonverbal stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1179–1214. https://doi.org/10.1037/xlm0000086

Craik, F. I. M., Eftekhari, E., & Binns, M. A. (2018). Effects of divided attention at encoding and retrieval: Further data. *Memory & Cognition*, *46*(8), 1263–1277. https://doi.org/10.3758/s13421-018-0835-3

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The Effects of

    Divided Attention on\nEncoding and Retrieval Processes in Human Memory. *Journal of*

    Experimental Psychology: General, *125*(2), 159–180. https://doi.org/10.1037/0096-

    3445.125.2.159

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context

    variability in cued recall. *Journal of Memory and Language*, *64*(2), 119–132.

    https://doi.org/10.1016/j.jml.2010.10.001

Cutler, R. A., Jeon, J., & Polyn, S. M. (2020, August). *Characterizing the interaction of*

    temporal and semantic information in categorized memory search [Poster session].

    Context and Episodic Memory Symposium. Philadelphia, PA.

Damaso, K., Williams, P., & Heathcote, A. (2020). Evidence for different types of errors being

    associated with different types of post-error changes. *Psychonomic Bulletin & Review*,

    27(3), 435–440. https://doi.org/10.3758/s13423-019-01675-w

Davis, E., Chemnitz, E., Collins, T. K., Geerligs, L., & Campbell, K. L. (2020). *Looking the*

    same, but remembering differently: Preserved eye-movement synchrony with age during

    movie-watching [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/xazdw

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects

    upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*,

    111(5), E618–E625. https://doi.org/10.1073/pnas.1317557111

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015).

    Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*,

    18(3), 470–475. https://doi.org/10.1038/nn.3940

deBettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2018). Forgetting from lapses of

    sustained attention. *Psychonomic Bulletin & Review*, *25*(2), 605–611.

    https://doi.org/10.3758/s13423-017-1309-5

Decker, A. L., Dubois, M., Duncan, K., & Finn, A. S. (2022). Pay attention and you might miss

    it: Greater learning during attentional lapses. *Psychonomic Bulletin & Review*.

    https://doi.org/10.3758/s13423-022-02226-6

Decker, A. L., & Duncan, K. (2020). Acetylcholine and the complex interdependence of memory

    and attention. *Current Opinion in Behavioral Sciences*, *32*, 21–28.

    https://doi.org/10.1016/j.cobeha.2020.01.013

Decker, A. L., Finn, A., & Duncan, K. (2020). Errors lead to transient impairments in memory

    formation. *Cognition*, *204*, 104338. https://doi.org/10.1016/j.cognition.2020.104338

Dester, A. M., Lazarus, L. T. T., Uitvlugt, M. G., & Healey, K. M. (2020, August). *A Test of*

    Retrieved Context Models: Dynamics of Recall After Incidental Encoding [Poster

    session]. Context and Episodic Memory Symposium. Philadelphia, PA.

Diamond, N. B., & Levine, B. (2020). Linking Detail to Temporal Structure in Naturalistic-

    Event Recall. *Psychological Science*, *31*(12), 1557–1572.

    https://doi.org/10.1177/0956797620958651

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the

    sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277.

    https://doi.org/10.1037/a0034024

DuBrow, S., & Davachi, L. (2014). Temporal Memory Is Shaped by Encoding Stability and

    Intervening Item Reactivation. *Journal of Neuroscience*, *34*(42), 13998–14005.

    https://doi.org/10.1523/JNEUROSCI.2535-14.2014

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*. https://doi.org/10.1016/j.nlm.2016.07.011

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141–146. https://doi.org/10.1016/j.cobeha.2017.08.003

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology (2006)*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: Episodic memory decisions induce lingering mnemonic biases. *Science (New York, N.Y.)*, *337*(6093), 485–487. https://doi.org/10.1126/science.1221936

Duncan, K., & Shohamy, D. (2016). Memory states influence value-based decisions. *Journal of Experimental Psychology. General*, *145*(11), 1420–1426. https://doi.org/10.1037/xge0000231

Dutilh, G., van Ravenzwaaij, D., Nieuwenhuis, S., van der Maas, H. L. J., Forstmann, B. U., & Wagenmakers, E.-J. (2012). How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology*, *56*(3), 208–216. https://doi.org/10.1016/j.jmp.2012.04.001

Earhard, M. (1972). *CUED RECALL AND FREE RECALL AS A FUNCTION OF THE NUMBER OF ITEMS PER CUE*. *26*(1).

Eichenbaum, H. (2017). On the Integration of Space, Time, and Memory. *Neuron*, *95*(5), 1007–1018. https://doi.org/10.1016/j.neuron.2017.06.036

Einhauser, W., Koch, C., & Carter, O. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00018

Elshiekh, A., & Rajah, M. N. (2021). Trial-by-trial fluctuations in post-stimulus attention during memory encoding predict subsequent associative context memory performance. *bioRxiv*, 2021.02.16.431454. https://doi.org/10.1101/2021.02.16.431454

Esterman, M., Liu, G., Okabe, H., Reagan, A., Thai, M., & DeGutis, J. (2015). Frontal eye field involvement in sustaining visual attention: Evidence from transcranial magnetic stimulation. *NeuroImage*, *111*, 542–548. https://doi.org/10.1016/j.neuroimage.2015.01.044

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Esterman, M., Poole, V., Liu, G., & DeGutis, J. (2017). Modulating Reward Induces Differential Neurocognitive Approaches to Sustained Attention. *Cerebral Cortex*, *27*(8), 4022–4032. https://doi.org/10.1093/cercor/bhw214

Esterman, M., Rosenberg, M. D., & Noonan, S. K. (2014). Intrinsic Fluctuations in Sustained Attention and Distractor Processing. *Journal of Neuroscience*, *34*(5), 1724–1730. https://doi.org/10.1523/JNEUROSCI.2658-13.2014

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/10.1016/j.copsyc.2019.03.005

Everitt, B. J., & Robbins, T. W. (1997). CENTRAL CHOLINERGIC SYSTEMS AND COGNITION. *Annual Review of Psychology*, *48*(Volume 48, 1997), 649–684. https://doi.org/10.1146/annurev.psych.48.1.649

Ezzyat, Y., & Davachi, L. (2010). What Constitutes an Episode in Episodic Memory?:

    Psychological Science. https://doi.org/10.1177/0956797610393742

Ezzyat, Y., & Davachi, L. (2014). Similarity Breeds Proximity: Pattern Similarity within and

    across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity.

    Neuron, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Fastenrath, M., Coynel, D., Spalek, K., Milnik, A., Gschwind, L., Roozendaal, B.,

    Papassotiropoulos, A., & Quervain, D. J. F. de. (2014). Dynamic Modulation of

    Amygdala–Hippocampal Connectivity by Emotional Arousal. *Journal of Neuroscience*,

    34(42), 13935–13947. https://doi.org/10.1523/JNEUROSCI.0786-14.2014

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using

    G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*,

    41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical

    power analysis program for the social, behavioral, and biomedical sciences. *Behavior*

    Research Methods, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and

    difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586–592.

    https://doi.org/10.3758/s13423-012-0367-y

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman,

    M. (2015). Sustained Attention Across the Life Span in a Sample of 10,000: Dissociating

    Ability and Strategy. *Psychological Science*, *26*(9), 1497–1510.

    https://doi.org/10.1177/0956797615594896

Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J., & Esterman, M. (2018). Tracking

behavioral and neural fluctuations during sustained attention: A robust replication and

extension. *NeuroImage*, *171*, 148–164. https://doi.org/10.1016/j.neuroimage.2018.01.002

Foucher, J. R., Otzenberger, H., & Gounot, D. (2004). Where arousal meets attention: A

simultaneous fMRI and EEG recording study. *NeuroImage*, *22*(2), 688–697.

https://doi.org/10.1016/j.neuroimage.2004.01.048

Garlitch, S. M., & Wahlheim, C. N. (2020). The role of attentional fluctuation during study in

recollecting episodic changes at test. *Memory & Cognition*, *48*(5), 800–814.

https://doi.org/10.3758/s13421-020-01018-4

Gelman, A., & Carlin, J. (2017). Some Natural Solutions to the p-Value Communication

Problem—And Why They Won't Work. *Journal of the American Statistical Association*,

112(519), 899–901. https://doi.org/10.1080/01621459.2017.1311263

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks

changes in control state predicted by the adaptive gain theory of locus coeruleus function.

Cognitive, Affective, & Behavioral Neuroscience, *10*(2), 252–269.

https://doi.org/10.3758/CABN.10.2.252

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind

wandering. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00031

Haring, J. H., & Davis, J. N. (1985). Differential distribution of locus coeruleus projections to

the hippocampal formation: Anatomical and biochemical evidence. *Brain Research*,

325(1), 366–369. https://doi.org/10.1016/0006-8993(85)90342-7

Harley, C. W. (2007). Norepinephrine and the dentate gyrus. In H. E. Scharfman (Ed.), *Progress in Brain Research* (Vol. 163, pp. 299–318). Elsevier. https://doi.org/10.1016/S0079-6123(07)63018-0

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527. https://doi.org/10.3758/s13428-015-0588-x

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*, 575–596. https://doi.org/10.1037/a0033715

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720. https://doi.org/10.3758/s13423-018-1537-3

Henríquez, R. A., Chica, A. B., Billeke, P., & Bartolomeo, P. (2016). Fluctuating Minds: Spontaneous Psychophysical Variability during Mind-Wandering. *PLOS ONE*, *11*(2), e0147174. https://doi.org/10.1371/journal.pone.0147174

Hershman, R., & Henik, A. (2019). Dissociation between reaction time and pupil dilation in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1899–1909. https://doi.org/10.1037/xlm0000690

Hess, E. H., & Polt, J. M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. *Science*, *132*(3423), 349–350. https://doi.org/10.1126/science.132.3423.349

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause
mnemonic trade-offs between local boundary processing and across-trial associative
binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7),
1075. https://doi.org/10.1037/xlm0000503

Hoffing, R., & Seitz, A. R. (2014). Pupillometry as a Glimpse into the Neurochemical Basis of
Human Memory Encoding. *Journal of Cognitive Neuroscience*, *27*(4), 765–774.
https://doi.org/10.1162/jocn_a_00749

Hong, L., He, H., & Sajda, P. (2023). Capturing Interactions between Arousal and Cortical
Dynamics with Simultaneous Pupillometry and EEG-fMRI. *2023 11th International
Conference on Affective Computing and Intelligent Interaction Workshops and Demos
(ACIIW)*, 1–8. https://doi.org/10.1109/ACIIW59127.2023.10388080

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free
recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(4),
923–941. https://doi.org/10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002a). A Distributed Representation of Temporal Context.
Journal of Mathematical Psychology, *46*(3), 269–299.
https://doi.org/10.1006/jmps.2001.1388

Howard, M. W., & Kahana, M. J. (2002b). When Does Semantic Similarity Help Episodic
Retrieval? *Journal of Memory and Language*, *46*(1), 85–98.
https://doi.org/10.1006/jmla.2001.2798

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *Journal
of Vision*, *11*(11), 1282. https://doi.org/10.1167/11.11.1282

Jayakumar, M., Balusu, C., & Aly, M. (2023). Attentional fluctuations and the temporal

organization of memory. *Cognition*, *235*, 105408.

https://doi.org/10.1016/j.cognition.2023.105408

Jeffreys, H. (1961). *The Theory of Probability*. OUP Oxford.

Jones, G. V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of

sequential position. *Journal of Experimental Psychology: General*, *105*(3), 277–293.

https://doi.org/10.1037/0096-3445.105.3.277

Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition.

Trends in Cognitive Sciences, *0*(0). https://doi.org/10.1016/j.tics.2020.03.005

Kafkas, A., & Montaldi, D. (2011). Recognition Memory Strength is Predicted by Pupillary

Responses at Encoding While Fixation Patterns Distinguish Recollection from

Familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989.

https://doi.org/10.1080/17470218.2011.588335

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1),

103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency

and lag recency effects in free recall. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, *28*(3), 530–540. https://doi.org/10.1037/0278-7393.28.3.530

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756),

1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kahneman, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task.

Perception & Psychophysics, *2*(3), 101–105. https://doi.org/10.3758/BF03210302

Kam, J. W. Y., & Handy, T. C. (2014). Differential recruitment of executive resources during

    mind wandering. *Consciousness and Cognition*, *26*, 51–63.

    https://doi.org/10.1016/j.concog.2014.03.002

Kane, M. J., & McVay, J. C. (2012). What Mind Wandering Reveals About Executive-Control

    Abilities and Failures. *Current Directions in Psychological Science*, *21*(5), 348–354.

    https://doi.org/10.1177/0963721412454875

Kang, O. E., Huffer, K. E., & Wheatley, T. P. (2014). Pupil Dilation Dynamics Track Attention

    to High-Level Information. *PLOS ONE*, *9*(8), e102463.

    https://doi.org/10.1371/journal.pone.0102463

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical*

    Association, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kaufman, A. M., Geiller, T., & Losonczy, A. (2020). A Role for the Locus Coeruleus in

    Hippocampal CA1 Place Cell Reorganization during Spatial Reward Learning. *Neuron*,

    105(6), 1018-1026.e4. https://doi.org/10.1016/j.neuron.2019.12.029

Kelly, A. J., & Beran, M. J. (2021). Multi-trial free recall dynamics in preschool children and

    young adults. *Journal of Cognitive Psychology*, *33*(8), 837–852.

    https://doi.org/10.1080/20445911.2021.1977815

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine

    release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and

    memory. *Proceedings of the National Academy of Sciences*, *113*(51), 14835–14840.

    https://doi.org/10.1073/pnas.1616515114

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*,

    330(6006), 932–932. https://doi.org/10.1126/science.1192439

Kozunova, G. L., Sayfulina, K. E., Prokofyev, A. O., Medvedev, V. A., Rytikova, A. M., Stroganova, T. A., & Chernyshev, B. V. (2022). Pupil dilation and response slowing distinguish deliberate explorative choices in the probabilistic learning task. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(5), 1108–1129. https://doi.org/10.3758/s13415-022-00996-z

Kragel, J. E., & Voss, J. L. (2021). Temporal context guides visual exploration during scene recognition. *Journal of Experimental Psychology: General*, *150*(5), 873–889. https://doi.org/10.1037/xge0000827

Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, *147*(8), 1111. https://doi.org/10.1037/xge0000411

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, *6*(3), 299–312. https://doi.org/10.1177/1745691611406925

Kucyi, A., Esterman, M., Riley, C. S., & Valera, E. M. (2016). Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences*, *113*(48), 13899–13904. https://doi.org/10.1073/pnas.1611743113

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral Cortex*, *27*(3), 1831–1840. https://doi.org/10.1093/cercor/bhw029

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

LaRocque, J. J., Eichenbaum, A. S., Starrett, M. J., Rose, N. S., Emrich, S. M., & Postle, B. R. (2015). The short- and long-term fates of memory items retained outside the focus of attention. *Memory & Cognition*, *43*(3), 453–468. https://doi.org/10.3758/s13421-014-0486-y

Larsen, R. S., & Waters, J. (2018). Neuromodulatory Correlates of Pupil Dilation. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00021

Lazarus, L. T. T., Dester, A. M., Uitvlugt, M. G., & Healey, K. M. (2020, August). *Modeling of Reduced Temporal Contiguity for Distinct Items* [Poster session]. Context and Episodic Memory Symposium. Philadelphia, PA.

Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, *52*(6), 362–375. https://doi.org/10.1016/j.jmp.2008.03.002

Long, N. M., & Kahana, M. J. (2017). Modulation of task demands suggests that semantic processing interferes with the formation of episodic associations. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(2), 167–176. https://doi.org/10.1037/xlm0000300

Lowenstein, O., Feinberg, R., & Loewenfeld, I. (1963). Pupillary Movements During Acute and Chronic Fatigue: A New Test for the Objective Evaluation of Tiredness. *Investigative Ophthalmology & Visual Science*, *2*(2), 138–157.

Madan, C. R., Fujiwara, E., Caplan, J. B., & Sommer, T. (2017). Emotional arousal impairs

    association-memory: Roles of amygdala and hippocampus. *NeuroImage*, *156*, 14–28.

    https://doi.org/10.1016/j.neuroimage.2017.04.065

Madore, K. P., Khazenzon, A. M., Backes, C. W., Jiang, J., Uncapher, M. R., Norcia, A. M., &

    Wagner, A. D. (2020). Memory failure predicted by attention lapsing and media

    multitasking. *Nature*, *587*(7832), Article 7832. https://doi.org/10.1038/s41586-020-2870-

    z

Makovac, E., Fagioli, S., Watson, D. R., Meeten, F., Smallwood, J., Critchley, H. D., &

    Ottaviani, C. (2019). Response time as a proxy of ongoing mental state: A combined

    fMRI and pupillometry study in Generalized Anxiety Disorder. *NeuroImage*.

    https://doi.org/10.1016/j.neuroimage.2019.02.038

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further

    investigations of sustained attention to response. *Neuropsychologia*, *37*(6), 661–670.

    https://doi.org/10.1016/S0028-3932(98)00127-4

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall.

    Memory, *20*(5), 511–517. https://doi.org/10.1080/09658211.2012.683010

Martarelli, C., & Ovalle-Fresa, R. (2021). *In sight, out of mind? Effects of task disengagement on

    visual recall precision*. PsyArXiv. https://doi.org/10.31234/osf.io/fdvzs

Mathot, S. (2018). *Pupillometry: Psychology, Physiology, and Function* (1). *1*(1), Article 1.

    https://doi.org/10.5334/joc.18

Mathôt, S., & Vilotijević, A. (2023). Methods in cognitive pupillometry: Design, preprocessing,

    and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077.

    https://doi.org/10.3758/s13428-022-01957-7

181

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 525. https://doi.org/10.1037/a0025896

Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, *11*, e70510. https://doi.org/10.7554/eLife.70510

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 978–984. https://doi.org/10.1037/xlm0000216

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLOS ONE*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, *17*(4), 510–515. https://doi.org/10.3758/PBR.17.4.510

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*, 406–419. https://doi.org/10.1037/a0024377

Mulligan, N. W., & Hartman, M. (1996). Divided attention and indirect memory tests. *Memory & Cognition*, *24*(4), 453–465. https://doi.org/10.3758/BF03200934

Murdock Jr., B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. https://doi.org/10.1037/h0045106

Murty, V. P., McKinney, R. A., DuBrow, S., Jalbrzikowski, M., Haas, G. L., & Luna, B. (2018).

    Differential patterns of contextual organization of memory in first-episode psychosis. *Npj*

    Schizophrenia, *4*(1), 1–6. https://doi.org/10.1038/s41537-018-0046-8

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and

    predicts later retrieval success for declarative memories of natural scenes. *Journal of*

    Vision, *13*(2), 11. https://doi.org/10.1167/13.2.11

Naveh-Benjamin, M., & Brubaker, M. S. (2019). Are the effects of divided attention on memory

    encoding processes due to the disruption of deep-level elaborative processes? Evidence

    from cued- and free-recall tasks. *Journal of Memory and Language*, *106*, 108–117.

    https://doi.org/10.1016/j.jml.2019.02.007

Naveh-Benjamin, M., Craik, F. I., Guez, J., & Dori, H. (1998). Effects of divided attention on

    encoding and retrieval processes in human memory: Further support for an asymmetry.

    Journal of Experimental Psychology: Learning, Memory, and Cognition, *24*(5), 1091.

Newman, E. L., Gupta, K., Climer, J. R., Monaghan, C. K., & Hasselmo, M. E. (2012).

    Cholinergic modulation of cognitive processing: Insights drawn from computational

    models. *Frontiers in Behavioral Neuroscience*, *6*.

    https://doi.org/10.3389/fnbeh.2012.00024

Nguyen, K., & McDaniel, M. A. (2015). The picture complexity effect: Another list composition

    paradox. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4),

    1026. https://doi.org/10.1037/xlm0000071

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-

    error slowing: An orienting account. *Cognition*, *111*(2), 275–279.

    https://doi.org/10.1016/j.cognition.2009.02.002

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, *42*(3), 306–317. https://doi.org/10.1111/j.1469-8986.2005.00291.x

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *117*(9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x

Palombo, D. J., Di Lascio, J. M., Howard, M. W., & Verfaellie, M. (2019). Medial Temporal Lobe Amnesia Is Associated with a Deficit in Recovering Temporal Context. *Journal of Cognitive Neuroscience*, *31*(2), 236–248. https://doi.org/10.1162/jocn_a_01344

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*(1), 56–64. https://doi.org/10.1016/j.ijpsycho.2011.10.002

Parasuraman, R. (2000). *The Attentive Brain*. MIT Press.

Parkin, A. J. (1981). Determinants of cued recall. *Current Psychology*, *1*(3), 291–300. https://doi.org/10.1007/BF03186739

Patton, P. E., & McNaughton, B. (1995). Connection matrix of the hippocampal formation: I. The dentate gyrus. *Hippocampus*, *5*(4), 245–286. https://doi.org/10.1002/hipo.450050402

Peris-Yague, A., Frank, D., & Strange, B. (2021, March). *Emotional and perceptual salience modulate conditional response probability curves in free recall* [Poster session]. Cognitive Neuroscience Society virtual meeting.

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*, 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, *12*(1), 24–30. https://doi.org/10.1016/j.tics.2007.10.010

Polyn, S. M., McCluey, J. D., Morton, N. W., Woolard, A. A., Luksik, A. S., & Heckers, S. (2015). Temporal context and the organisational impairment of memory search in schizophrenia. *Cognitive Neuropsychiatry*, *20*(4), 296–310. https://doi.org/10.1080/13546805.2015.1031372

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129. https://doi.org/10.1037/a0014420

Poskanzer, C., & Aly, M. (2023). Switching between External and Internal Attention in Hippocampal Networks. *Journal of Neuroscience*, *43*(38), 6538–6552. https://doi.org/10.1523/JNEUROSCI.0029-23.2023

Purcell, B. A., & Kiani, R. (2016). Neural Mechanisms of Post-error Adjustments of Decision Policy in Parietal Cortex. *Neuron*, *89*(3), 658–671. https://doi.org/10.1016/j.neuron.2015.12.027

Reinhard, G., Lachnit, H., & König, S. (2007). Effects of stimulus probability on pupillary dilation and reaction time in categorization. *Psychophysiology*, *44*(3), 469–475. https://doi.org/10.1111/j.1469-8986.2007.00512.x

Richer, F., & Beatty, J. (1987). Contrasting Effects of Response Uncertainty on the Task-Evoked

Pupillary Response and Reaction Time. *Psychophysiology*, *24*(3), 258–262.

https://doi.org/10.1111/j.1469-8986.1987.tb00291.x

Rijn, H. van, Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil Dilation Co-Varies

with Memory Strength of Individual Traces in a Delayed Response Paired-Associate

Task. *PLOS ONE*, *7*(12), e51134. https://doi.org/10.1371/journal.pone.0051134

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday

Attention: Variation in Mind Wandering and Memory in a Lecture. *Applied Cognitive

Psychology*, *26*(2), 234–242. https://doi.org/10.1002/acp.1814

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!':

Performance correlates of everyday attentional failures in traumatic brain injured and

normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-

3932(97)00015-8

Rosenberg, M. D., Finn, E. S., Constable, R. T., & Chun, M. M. (2015). Predicting moment-to-

moment attentional state. *NeuroImage*, *114*, 249–256.

https://doi.org/10.1016/j.neuroimage.2015.03.032

Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017).

Characterizing Attention with Predictive Network Models. *Trends in Cognitive Sciences*,

21(4), 290–302. https://doi.org/10.1016/j.tics.2017.01.011

Rosenberg, M. D., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention

in the face of distraction: A novel gradual onset continuous performance task. *Journal of

Vision*, *11*(11), 127–127. https://doi.org/10.1167/11.11.127

Rothlein, D., DeGutis, J., & Esterman, M. (2018). Attentional Fluctuations Influence the Neural Fidelity and Connectivity of Stimulus Representations. *Journal of Cognitive Neuroscience*, *30*(9), 1209–1228. https://doi.org/10.1162/jocn_a_01306

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, *18*(4), 682–689. https://doi.org/10.3758/s13423-011-0088-7

Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010). *Psychological Bulletin*, *139*, 241–247. https://doi.org/10.1037/a0029008

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ruivo, L. M. T.-G., Baker, K. L., Conway, M. W., Kinsley, P. J., Gilmour, G., Phillips, K. G., Isaac, J. T. R., Lowry, J. P., & Mellor, J. R. (2017). Coordinated Acetylcholine Release in Prefrontal Cortex and Hippocampus Is Associated with Arousal and Reward on Distinct Timescales. *Cell Reports*, *18*(4), 905–917. https://doi.org/10.1016/j.celrep.2016.12.085

Ruiz, N. A., Thieu, M. K., & Aly, M. (2021). Cholinergic modulation of hippocampally mediated attention and perception. *Behavioral Neuroscience*, *135*(1), 51–70. https://doi.org/10.1037/bne0000405

Samuels, E. R., & Szabadi, E. (2008). Functional Neuroanatomy of the Noradrenergic Locus Coeruleus: Its Roles in the Regulation of Arousal and Autonomic Function Part II: Physiological and Pharmacological Manipulations and Pathological Alterations of Locus Coeruleus Activity in Humans. *Current Neuropharmacology*, *6*(3), 254–285.

Santa, J. L., Ruskin, A. B., Snuttjer, D., & Baker, L. (1975). Retrieval in cued recall. *Memory & Cognition*, *3*(3), 341–348. https://doi.org/10.3758/BF03212922

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews*, *35*(2), 146–160. https://doi.org/10.1016/S0165-0173(01)00044-3

Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, *16*(11), 898–904. https://doi.org/10.1111/j.1467-9280.2005.01634.x

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689–699. https://doi.org/10.3758/MC.38.6.689

Semba, K. (2000). Multiple output pathways of the basal forebrain: Organization, chemical heterogeneity, and roles in vigilance. *Behavioural Brain Research*, *115*(2), 117–141. https://doi.org/10.1016/S0166-4328(00)00254-0

Seymour, T., Baker, C. A., & Gaunt, J. T. (2013). Combining Blink, Pupil, and Response Time Measures in a Concealed Knowledge Test. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00614

Siegelman, M. (2019). *Simulate an experiment with optimal stimuli segmentation*. GitHub. https://github.com/msieg/OptSeg_Reproducible

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). *False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant* (p. 555). American Psychological Association. https://doi.org/10.1037/14805-033

Simonsohn, U. (2014). *Posterior-Hacking: Selective Reporting Invalidates Bayesian Results Also* (SSRN Scholarly Paper 2374040). https://doi.org/10.2139/ssrn.2374040

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Smallwood, J., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition*, *12*(3), 452–484. https://doi.org/10.1016/S1053-8100(03)00018-7

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). *Going AWOL in the Brain: Mind Wandering Reduces Cortical Analysis of External Events*. *20*(3), 12.

Smallwood, J., & Schooler, J. W. (2006). *The restless mind.* Psychological Bulletin; US: American Psychological Association. https://doi.org/10.1037/0033-2909.132.6.946

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532–1539. https://doi.org/10.1037/a0024852

Steinmetz, K. R., Knight, A. G., & Kensinger, E. A. (2016). Neutral details associated with emotional events are encoded: Evidence from a cued recall paradigm. *Cognition and Emotion*, *30*(7), 1352–1360. https://doi.org/10.1080/02699931.2015.1059317

Strauch, C., Wang, C.-A., Einhäuser, W., Stigchel, S. V. der, & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*(8), 635–647. https://doi.org/10.1016/j.tins.2022.05.003

Tarder-Stoll, H., Jayakumar, M., Dimsdale-Zucker, H. R., Günseli, E., & Aly, M. (2020).

    Dynamic internal states shape memory retrieval. *Neuropsychologia*, *138*, 107328.

    https://doi.org/10.1016/j.neuropsychologia.2019.107328

Troyer, A. K., & Craik, F. I. M. (2000). The effect of divided attention on memory for items and

    their context. *Canadian Journal of Experimental Psychology/Revue Canadienne de*

    Psychologie Expérimentale, *54*(3), 161–171. https://doi.org/10.1037/h0087338

Troyer, A. K., Winocur, G., Craik, F. I. M., & Moscovitch, M. (1999). Source memory and

    divided attention: Reciprocal costs to primary and secondary tasks. *Neuropsychology*,

    13(4), 467–474. https://doi.org/10.1037/0894-4105.13.4.467

Turk-Browne, N. B., Golomb, J. D., & Chun, M. M. (2013). Complementary attentional

    components of successful memory encoding. *NeuroImage*, *66*, 553–562.

    https://doi.org/10.1016/j.neuroimage.2012.10.053

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal Proximity Links Unrelated News Events in

    Memory. *Psychological Science*, *30*(1), 92–104.

    https://doi.org/10.1177/0956797618808474

Uncapher, M. R., Hutchinson, J. B., & Wagner, A. D. (2011). Dissociable Effects of Top-Down

    and Bottom-Up Attention during Episodic Encoding. *Journal of Neuroscience*, *31*(35),

    12613–12628. https://doi.org/10.1523/JNEUROSCI.0152-11.2011

Uncapher, M. R., & Rugg, M. D. (2005). Effects of Divided Attention on fMRI Correlates of

    Memory Encoding. *Journal of Cognitive Neuroscience*, *17*(12), 1923–1935. Journal of

    Cognitive Neuroscience. https://doi.org/10.1162/089892905775008616

Uncapher, M. R., & Rugg, M. D. (2009). Selecting for memory? The influence of selective
attention on the mnemonic binding of contextual information. *Journal of Neuroscience*,
29(25), 8270–8279.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Focusing the search: Proactive and
retroactive interference and the dynamics of free recall. *Journal of Experimental
Psychology: Learning, Memory, and Cognition*, *39*(6), 1742–1756.
https://doi.org/10.1037/a0033743

Unsworth, N., & Miller, A. L. (2021). Individual Differences in the Intensity and Consistency of
Attention. *Current Directions in Psychological Science*, *30*(5), 391–400.
https://doi.org/10.1177/09637214211030266

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Individual differences in lapses of
sustained attention: Ocolumetric indicators of intrinsic alertness. *Journal of Experimental
Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000734

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention.
Cognitive, Affective, & Behavioral Neuroscience, *16*(4), 601–615.
https://doi.org/10.3758/s13415-016-0417-4

Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with
pupillometry. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(4), 638–664.
https://doi.org/10.3758/s13415-018-0594-4

Unsworth, N., Robison, M. K., & Miller, A. L. (2018). Pupillary correlates of fluctuations in
sustained attention. *Journal of Cognitive Neuroscience*, *30*(9), 1241–1253.
https://doi.org/10.1162/jocn_a_01251

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637

van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil Diameter Tracks Lapses of Attention. *PLOS ONE*, *11*(10), e0165274. https://doi.org/10.1371/journal.pone.0165274

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, *115*(2), E310–E316. https://doi.org/10.1073/pnas.1714082115

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426–432. https://doi.org/10.1037/a0022790

Wainstein, G., Rojas-Líbano, D., Crossley, N. A., Carrasco, X., Aboitiz, F., & Ossandón, T. (2017). Pupil Size Tracks Attentional Performance In Attention-Deficit/Hyperactivity Disorder. *Scientific Reports*, *7*(1), 8228. https://doi.org/10.1038/s41598-017-08246-w

Wakeland-Hart, C. D., Cao, S. A., deBettencourt, M. T., Bainbridge, W. A., & Rosenberg, M. D. (2022). Predicting visual memory across images and within individuals. *Cognition*, *227*, 105201. https://doi.org/10.1016/j.cognition.2022.105201

Wang, C.-A., Baird, T., Huang, J., Coutinho, J. D., Brien, D. C., & Munoz, D. P. (2018). Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Frontiers in Neurology*, *9*. https://doi.org/10.3389/fneur.2018.01029

Wang, Y. C., Adcock, R. A., & Egner, T. (2023). Toward an integrative account of internal and external determinants of event segmentation. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02375-2

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*, *9*(7), Article 7. https://doi.org/10.1038/nn1727

Whitlock, J., Hubbard, R., Ding, H., & Sahakyan, L. (2023). Trial-level fluctuations in pupil dilation at encoding reflect strength of relational binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001286

Williams, P., Heathcote, A., Nesbitt, K., & Eidels, A. (2016). Post-error recklessness and the hot hand. *Judgment and Decision Making*, *11*(2), 11.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*, 2331216518800869. https://doi.org/10.1177/2331216518800869

Xu, J., Friedman, D., & Metcalfe, J. (2018). Attenuation of Deep Semantic Processing during Mind Wandering: An ERP study. *Neuroreport*, *29*(5), 380–384. https://doi.org/10.1097/WNR.0000000000000978

Xu, J., & Metcalfe, J. (2016). Studying in the region of proximal learning reduces mind wandering. *Memory & Cognition*, *44*(5), 681–695. https://doi.org/10.3758/s13421-016-0589-8

Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M., Germine, L., Wilmer, J., DeGutis, J., & Esterman, M. (2021). Variable rather than extreme slow reaction times distinguish brain

states during sustained attention. *Scientific Reports*, *11*(1), Article 1.

https://doi.org/10.1038/s41598-021-94161-0

Yanko, M. R., & Spalek, T. M. (2013). Route familiarity breeds inattention: A driving simulator

study. *Accident Analysis & Prevention*, *57*, 80–86.

https://doi.org/10.1016/j.aap.2013.04.003

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011).

Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus

activity in nondemented older adults. *Hippocampus*, *21*(9), 968–979.

https://doi.org/10.1002/hipo.20808

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in*

Neurosciences, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yi, D.-J., & Chun, M. M. (2005). Attentional Modulation of Learning-Related Repetition

Attenuation Effects in Human Parahippocampal Cortex. *Journal of Neuroscience*, *25*(14),

3593–3600. https://doi.org/10.1523/JNEUROSCI.4677-04.2005

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event

perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273.

https://doi.org/10.1037/0033-2909.133.2.273

Zhang, F., & Emberson, L. L. (2019). Opposing Timing Constraints Severely Limit the Use of

Pupillometry to Investigate Visual Statistical Learning. *Frontiers in Psychology*, *10*.

https://www.frontiersin.org/article/10.3389/fpsyg.2019.01792

Zhang, Q., Griffiths, T., & Norman, K. (2021). *Optimal policies for free recall*. PsyArXiv.

https://doi.org/10.31234/osf.io/sgepb

Zhang, W., Zhang, Y., & Xu, J. (2020). Performance and Eye Metrics Correlates to Out-of-the-

    zone Sustained Attention in GradCPT. *Proceedings of the Human Factors and*

    Ergonomics Society Annual Meeting, *64*(1), 112–116.

    https://doi.org/10.1177/1071181320641029

Appendix A: Supplementary materials for Chapter 1

		Study 1	Study 2	Study 3	Study 4
lag-CRP	Lag (1 to 29)	Nearby > far away $F_{10,1640.60} = 29.63$ $p < 0.0001$ $\eta_p^2 = 0.32$ BF < 1/150	Nearby > far away $F_{11.69,783.11} = 14.89$ $p < 0.0001$ $\eta_p^2 = 0.18$ BF < 1/150	Nearby > far away $F_{11.57, 775.09} = 11.98$ $p < 0.0001$ $\eta_p^2 = 0.15$ BF < 1/150	Nearby > far away $F_{15.18,1866.65} = 10.06$ $p < 0.0001$ $\eta_p^2 = 0.08$ BF < 1/150
	Direction (Forward vs. Backward)	$F_{1,64} = 3.01$ $p = 0.08$ $\eta_p^2 = 0.05$ BF = 8.95	Forward > backward $F_{1,67} = 11.75$ $p = 0.001$ $\eta_p^2 = 0.15$ BF = 0.10	$F_{1,67} = 3.09$ $p = 0.08$ $\eta_p^2 = 0.04$ BF = 7.85	$F_{1,123} = 0.46$ $p = 0.50$ $\eta_p^2 = 0.004$ BF = 43.57
	Attentional State (In vs. Out)	$F_{1,64} = 0.86$ $p = 0.36$ $\eta_p^2 = 0.013$ BF = 36.84	$F_{1,67} = 0.98$ $p = 0.33$ $\eta_p^2 = 0.014$ BF = 29.83	$F_{1,67} = 0.30$ $p = 0.59$ $\eta_p^2 = 0.005$ BF = 36.96	$F_{1,123} = 1.25$ $p = 0.27$ $\eta_p^2 = 0.01$ BF = 36.20
	Lag x Direction	Forward > backward for nearby > far away $F_{11.53,737.65} = 2.59$ $p = 0.003$ $\eta_p^2 = 0.04$ BF = 0.017	Forward > backward for nearby > far away $F_{14.11,945} = 2.50$ $p = 0.002$ $\eta_p^2 = 0.036$ BF = 0.23	Forward > backward for nearby > far away $F_{16.07,1076.46} = 1.74$ $p = 0.04$ $\eta_p^2 = 0.025$ BF >150	$F_{14.91,1834.13} = 1.38$ $p = 0.15$ $\eta_p^2 = 0.01$ BF > 150
	Lag x Attentional State	$F_{13.89,888.83} = 0.70$ $p = 0.78$ $\eta_p^2 = 0.01$ BF > 150	$F_{15.39,1031.14} = 0.92$ $p = 0.55$ $\eta_p^2 = 0.014$ BF > 150	$F_{16.15,1081.96} = 1.09$ $p = 0.36$ $\eta_p^2 = 0.016$ BF > 150	$F_{17.46,2147.86} = 0.77$ $p = 0.73$ $\eta_p^2 = 0.006$ BF > 150
	Direction x Attentional State	$F_{1,64} = 2.33$ $p = 0.13$ $\eta_p^2 = 0.04$ BF = 10.97	$F_{1,67} = 0.05$ $p = 0.83$ $\eta_p^2 = 0.0008$ BF = 27.81	$F_{1,67} = 0.15$ $p = 0.70$ $\eta_p^2 = 0.002$ BF = 26.36	$F_{1,123} = 0.67$ $p = 0.42$ $\eta_p^2 = 0.005$ BF = 30.49
	Lag x Direction x Attentional State	$F_{12.52,801.24} = 0.91$ $p = 0.54$ $\eta_p^2 = 0.014$ BF > 150	$F_{14.89,997.60} = 1.31$ $p = 0.19$ $\eta_p^2 = 0.02$ BF > 150	$F_{16.23,1087.49} = 1.05$ $p = 0.39$ $\eta_p^2 = 0.015$ BF > 150	$F_{17.20,2115.97} = 0.75$ $p = 0.76$ $\eta_p^2 = 0.006$ BF > 150
lag-CRP +/- 1	Lag (+1 vs. -1)	+1 > -1 $F_{1,64} = 9.51$ $p = 0.003$ $\eta_p^2 = 0.13$ BF = 0.02	+1 > -1 $F_{1,67} = 13.11$ $p = 0.0006$ $\eta_p^2 = 0.16$ BF = 0.009	+1 > -1 $F_{1,67} = 6.48$ $p = 0.01$ $\eta_p^2 = 0.09$ BF = 0.23	+1 > -1 $F_{1,123} = 4.69$ $p = 0.03$ $\eta_p^2 = 0.04$ BF = 0.48
	Attentional State (In vs. Out)	$F_{1,64} = 0.90$ $p = 0.35$ $\eta_p^2 = 0.014$ BF = 5.42	$F_{1,67} = 0.32$ $p = 0.57$ $\eta_p^2 = 0.005$ BF = 6.58	$F_{1,67} = 0.15$ $p = 0.70$ $\eta_p^2 = 0.002$ BF = 6.99	$F_{1,123} = 1.37$ $p = 0.24$ $\eta_p^2 = 0.01$ BF = 5.70
	Lag x Attentional State	$F_{1,64} = 0.14$ $p = 0.71$	$F_{1,67} = 0.85$ $p = 0.36$	$F_{1,67} = 1.47$ $p = 0.23$	$F_{1,123} = 0.19$ $p = 0.66$

		Study 1	Study 2	Study 3	Study 4
		$\eta_p^2 = 0.002$ BF = 4.99	$\eta_p^2 = 0.013$ BF = 3.51	$\eta_p^2 = 0.02$ BF = 2.94	$\eta_p^2 = 0.002$ BF = 6.18
Event Analysis	Transition Type	$F_{1.06, 67.57} = 26.37$ $p < 0.0001$ $\eta_p^2 = 0.29$ BF < 1/150	$F_{1.07, 71.55} = 45.59$ $p < 0.001$ $\eta_p^2 = 0.41$ BF < 1/150	$F_{1.06, 70.80} = 26.49$ $p < 0.0001$ $\eta_p^2 = 0.28$ BF < 1/150	$F_{1.04, 127.75} = 14.77$ $p = 0.0002$ $\eta_p^2 = 0.11$ BF < 1/150
	Attentional State (In vs. Out)	$F_{1,64} = 1.45$ $p = 0.23$ $\eta_p^2 = 0.02$ BF = 7.42	$F_{1,67} = 0.51$ $p = 0.48$ $\eta_p^2 = 0.008$ BF = 7.89	$F_{1,67} = 1.39$ $p = 0.24$ $\eta_p^2 = 0.02$ BF = 7.31	$F_{1,123} = 0.76$ $p = 0.39$ $\eta_p^2 = 0.006$ BF = 9.49
	Transition Type x Attentional State	$F_{1.33, 85.06} = 0.75$ $p = 0.43$ $\eta_p^2 = 0.012$ BF = 14.21	$F_{1.13, 75.40} = 0.64$ $p = 0.45$ $\eta_p^2 = 0.01$ BF = 14.21	$F_{1.24, 82.82} = 1.30$ $p = 0.27$ $\eta_p^2 = 0.02$ BF = 11.31	$F_{1.08, 132.69} = 0.78$ $p = 0.39$ $\eta_p^2 = 0.006$ BF = 18.72

Supplementary Table 4. Summary of results across Studies 1-4 for the "state-specific denominator approach". Green shading indicates statistically significant results ($p < 0.05$) and gray shading indicates results that are not statistically significant. In = "in the zone" attentional state; Out = "out of the zone" attentional state, defined with the VTC analysis.