THE UNIVERSITY OF CHICAGO

# SUSTAINED ATTENTION'S FLOODLIGHT DYNAMICALLY INFLUENCES COGNITIVE AND NEURAL REPRESENTATIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

ANNA CORRIVEAU

CHICAGO, ILLINOIS

AUGUST 2026

ORCID iD: https://orcid.org/0000-0003-3122-7198

## Contents

## List of Figures

## List of Tables

## Acknowledgements

Thank you to my advisor Dr. Monica Rosenberg for your thoughtful mentorship, not only in your creative and detailed approach to research, but also in your consideration of your students as humans outside of the lab. You are the role model for what a mentor should be and if I am ever fortunate enough to advise students of my own, I will do my best to emulate you.

Thank you to my committee, Dr. Marc G. Berman, Dr. Edward K. Vogel, and Dr. Megan T. deBettencourt, for your guidance throughout graduate school and in navigating my next steps. I am so fortunate to have worked with and learned from such kind, talented researchers.

Thank you to Kristi Schonwald, Meyosha Smiley, and Ray Weathers for keeping the department running and knowing the answer to every single question.

Thank you to the CAB Lab members past and present---Ziwei, Nia, Isabel, Hayoung, Jin, Matthieu, Taylor, Omid, Shan, Anthony, Isaac, Louisa, Lucy, Dongfang, and many more amazing researchers---as well as the larger BLRB group for fostering an environment of support, curiosity, and fun.

Thank you to my past advisors Dr. Chris Baker and Dr. C. Shawn Green for teaching me the skills I needed to succeed in graduate school and for showing me that research can be fun.

Thank you to friends from past labs, including Shruti, Sanika, R'ay, Noah, Martin, Hunter, Lina, Adam, Susan, and Wan, for being friends through one of life's most difficult times (grad school apps) and beyond.

Thank you to the friends I have made during graduate school, including Alfred, Anita, Claire, Elizabeth, Ezra, Geneva, Grant, Jack, John, Layn, Louisa, Madeline, Nakwon, Pinar, Trevor, Sabina, Sunny, and Susanne, for many movie/reality TV/card nights, intramural games, and for generally keeping spirits high.

Thank you to Alexis for being my academic buddy and for always being ready for a 2-hour vent session or to read a rough draft of any document. I cannot wait to keep tackling life's milestones with you.

Thank you to Jenny for giving the best advice and just being a great friend always.

Thank you to Claire, Brenna, and Ben for being my life outside of work. I love you guys.

Thank you to my family, and especially to my parents, for finding the science museums on every family vacation and for teaching me how to be a scientist from a young age.

Finally, thank you to Henry for being my best friend and the best partner I could imagine. You make every part of life better.

# Abstract

Despite our best efforts, humans are unable to maintain the focus of our attention indefinitely. The mnemonic consequences of attentional lapses are evident in our daily lives, such as forgetting where we left our phone after mindlessly setting it down. Our ability to sustain attention is consistent at the trait level---individuals attend either reliably well or reliably poorly---but varies at the state level, with all individuals fluctuating between periods of high and low sustained attentional states across seconds to minutes. Yet, we lack a full understanding of the brain-based mechanisms involved in attention's dynamics. The current dissertation aims to better characterize the nature of sustained attention through investigation of its neural and behavioral correlates. Using patterns of brain connectivity, I identify distributed networks that track sustained attention across individuals (Chapter 1), across contexts (Chapter 2), and across moments (Chapter 3), suggesting that attentional maintenance reconfigures functional relationships throughout the brain. I then explore the consequences of dynamic attentional states on the fate of information reflected in memory (Chapter 4) and in neural representations during encoding (Chapters 5 and 6). Evidence suggests that heightened sustained attentional states, rather than sharpening spotlight-like selectivity, increase processing capacity broadly in a manner more akin to a floodlight. Increased processing is further supported by a strengthening in representational fidelity under better attentional states as measured with spatially- and temporally-resolved neural signals. Results converge on the characterization of sustained attention as a global mechanism that supports cognition through the strengthening of functional connections and information representations in the brain.

# Introduction

As humans, there is an intuitive sense that our ability to pay attention waxes and wanes over time—from seconds to minutes to hours. Yet, cognitive neuroscience lacks an understanding of the cognitive and neural mechanisms that underlie the ability to maintain one's attention on a relevant task or goal, termed sustained attention. In this thesis, I probe the nature of sustained attention by examining its behavioral and neural basis across people, time, and contexts. Further, I seek to understand the nature of sustained attention by investigating its interactions with memory with which it is tightly linked. Ultimately, the studies included in this thesis characterize sustained attention as a global mechanism that shapes processing and memory through changes in functional organization and representations in the brain.

A core feature of sustained attention is that it varies from moment to moment, leading to periods of both engaged attention as well as attentional lapses. One means of quantifying fluctuations in sustained attention over time as well as across individuals is through the use of continuous performance tasks (CPTs), which require frequent responding. While task-specific features of CPTs may change from one study to the next, their core structure is the same: Participants respond to frequent-category stimuli (90%) while withholding responses to infrequent stimuli (10%). Regular responding leads participants to a pattern of mindless pressing such that, under sustained attentional lapses, participants may fail to inhibit a response to a rare item. Moreover, responses to frequent trials can be analyzed to monitor continuous sustained attentional dynamics. Response time speed and variability during CPTs reliably predict upcoming lapses in sustained attention, such that lapses tend to be preceded by faster and more variable pressing (Esterman et al.; 2013; Rosenberg et al., 2013; deBettencourt et al., 2018; deBettencourt et al., 2019). As such, CPTs provide a standardized, highly-controlled means of tracking fluctuations in

attention over time. Despite their utility in providing objective indices of sustained attentional state, however, CPTs do not closely mirror everyday life. Instead, our naturalistic experiences tend to evolve continuously over time and often involve complex, multimodal, and narrative components. To examine more naturalistic fluctuations in sustained attention, we ask participants to report subjective changes in engagement while they watch movies and listen to stories. These narrative stimuli provide a less-controlled task set, allowing participants to experience them as they would in daily life. Through the use of both controlled and naturalistic task paradigms, we probe features of sustained attention that generalize broadly across contexts.

An additional means of understanding sustained attention is by characterizing its neural basis. A growing body of work leverages functional magnetic resonance imaging (fMRI) connectivity, a measure of dependence between brain regions which is largely stable across individuals (Finn et al., 2015; Gratton et al., 2018; Miranda-Dominguez et al., 2014) but demonstrates characteristic changes as a result of tasks (Gratton et al., 2016; Greene et al., 2020; Lynch et al., 2016; Shine et al., 2020). For example, functional connectivity-based signatures reliably predict sustained attention ability across individuals (Kardan et al., 2022; Rosenberg et al., 2016; Rosenberg et al., 2018; Wu et al., 2020; Yoo et al., 2022) as well as fluctuations within individuals (Esterman et al., 2013; Jones et a., 2024; Kucyi et al., 2017; Rosenberg et al., 2020; Seeburger et al., 2024). Additionally, neural representations (stimulus-related patterns of brain activity) are also more consistent across individuals under better attentional states (Rothlein et al., 2018), suggesting that attention strengthens the reliability of neural activity. Taken together, fluctuations in sustained attention reconfigure patterns of brain activity and connectivity which may underlie observable changes in behavior.

In addition to identifying behavioral and neural signatures of attentional fluctuations, it is important to understand the consequences of these fluctuations for how information is processed and remembered. Previous work demonstrates that high attentional state, quantified by response time signatures, not only leads to better sustained attention task performance but also increased memory later on (deBettencourt et al., 2018, Decker et al., 2023a, Decker et al., 2023b; Wakeland-Hart et al., 2022). Further, neural models of sustained attention generalize outside of controlled task contexts to predict narrative engagement and memory during naturalistic stimulus viewing (Song et al., 2021). Yet, it is not clear how far the impacts of sustained attention extend. Are moments of increased sustained attention characterized by increased processing for all information or relevant information selectively enhanced under better attention? What neural mechanisms support attention-related changes in memory?

The following chapters characterize sustained attention by examining its reliable neural and behavioral signatures across individuals, moments, and contexts. Further, I explore the link between sustained attention and memory to reveal the complex relationship between these processes. The first chapter investigates features of whole-brain patterns of functional connectivity, or connectomes, which predict sustained attention performance across individuals. I show how these features can be used in combination with supervised connectivity-based models to improve the generalization of predictive models across samples. Chapter 2 demonstrates robust, modality-general prediction of sustained attention from whole-brain patterns of functional connectivity, suggesting that connectivity-based models of sustained attention identify modality-agnostic features which can be used to predict performance regardless of task-modality. In Chapter 3 I leverage a high-frequency measure of functional connectivity to investigate high-frequency fluctuations in visual and auditory sustained attention which occur over a matter of seconds.

Chapter 4 includes behavioral studies which explore how fluctuations in sustained attention impact memory. In particular, these studies investigate how sustained attentional state at encoding affects memory for both task-relevant and task-irrelevant stimuli. In Chapter 5, I test the extent to which fluctuations in subjective engagement impact the fidelity of stimulus-relevant representations in fMRI signals. Finally, Chapter 6 explores the availability of stimulus information in high-frequency neural signals. Results converge to characterize sustained attention as a global, modality-general mechanism supporting cognition through the strengthening of functional networks and information representations in the brain.

# Chapter 1: Functional connectome stability is a marker of cognitive performance

Although essential for daily life, individual abilities vary for cognitive processes such as sustaining attention across time and holding items in working memory. Identifying brain-based markers of such abilities has been a goal of much recent work in cognitive neuroscience. Brain-based predictive modeling may provide insight both into mechanisms broadly underlying cognitive ability as well as the ways in which individuals meaningfully differ.

Network neuroscience posits that cognitive processes arise from large-scale interactions between distributed sets of brain regions, and work characterizing these interactions has found evidence that they relate to behavior (for a review, see Bassett and Sporns 2017). In particular, network neuroscientific approaches have demonstrated that whole-brain fMRI functional connectivity predicts individual differences in behavioral phenotypes using methods such as connectome-based predictive modeling (CPM; Rosenberg et al., 2016; Shen et al., 2017). In CPM, a set of functional connections between brain regions are identified to serve as model features for a particular predicted phenotype. This method has successfully been implemented to predict cognitive processes and behavioral measures, including sustained attention, fluid intelligence, processing speed, and working memory (Finn et al., 2015; Rosenberg et al., 2016; Avery et al., 2020; Gao et al., 2020; Yoo et al., 2022a). The growing list of CPM's predictive applications supports functional connectivity's utility as a predictive tool for a diverse set of outcome variables.

In addition to these supervised network-based predictive models, models based on individual- or group-related summary features of individuals' whole-brain patterns of functional connectivity, or connectomes, may predict aspects of cognitive and attentional function. One approach to quantifying connectome features is characterizing their network properties using graph

theory measures. These network properties, which characterize interactions between brain regions, have been related to variation in cognitive abilities such as general intelligence (van den Heuvel et al., 2009; Hilger et al., 2017) and working memory (Stanley et al., 2015), providing further support that differences in structural and functional brain organization underlie behavioral and cognitive performance. Here, we take a complementary approach, investigating whether similarities—or differences— in task-based network patterns within and across individuals relate to cognitive ability. Specifically, we define features of the whole-brain connectome that may meaningfully vary across individuals: connectome similarity to oneself (stability), connectome similarity to others (typicality and optimality), and their relationship with each other.

Connectome stability reflects the similarity of one's own functional connectivity pattern over time. Despite evidence that the connectome is stable enough to function as a "fingerprint" (Miranda-Dominguez et al., 2014; Finn et al., 2015; Gratton et al., 2018), there is some variability in stability across individuals which may be useful for prediction. Previously, resting state stability, measured as the similarity of an individual's connectome or network connectivity across repeated fMRI runs, was related to cognitive and social function in adolescents and young adults (Vanderwal et al., 2021; Fu et al., 2022), such that higher connectome stability was associated with better cognitive performance and social ability. Furthermore, connectome stability increases over the course of development but is reduced for individuals with psychiatric disorders (Kaufmann et al., 2017; Vanderwal et al., 2021; Fu et al., 2022). As functional connectivity patterns dynamically shift to reflect changes in cognitive state in response to task (Gonzalez-Castillo et al., 2015; Xie et al., 2018), one possibility is that connectome stability reflects the variability of an individual's cognitive states over time, such that more stable connectivity patterns between repeated resting-state or task runs may reflect a more on-task cognitive state which results in higher task

performance. Additionally, stability (i.e., similarity) between resting-state and task connectomes may reflect less task-specific reconfiguration which has previously been related to cognitive function (Schultz and Cole 2016). Here, we examine whether stability across task runs or between task types predicts cognitive performance.

Connectome typicality measures the similarity of an individual's connectome to the average connectome of others in a group. For the current study, we hypothesized two potential ways in which connectome typicality might relate to cognitive performance. First, previous work has shown that activity in largescale brain networks (e.g., the frontoparietal network) relates to attention and working memory performance in children and young adults (Satterthwaite et al., 2013; Rosenberg et al., 2020). Therefore, to the extent that a similar task-related pattern of connectivity is captured in the group-average, the degree to which an individual resembles others may be indicative of task-related processing and predictive of performance. Indeed, connectome typicality during movie watching was predictive of social function in development (Vanderwal et al., 2021).

This hypothesis about how connectome typicality relates to performance assumes that the group-average task connectome captures a meaningful "template" pattern of task-specific functional connectivity. However, it is also possible that patterns related to high task performance get lost in the group average. Instead, perhaps, a better proxy of a task-optimal template of connectivity is the average connectome of individuals who perform best on a given task. To investigate this possibility, we calculated each participant's connectome "optimality," which reflects the similarity of one's average task connectome to the average connectome of the highest task performers. A relationship between task connectome optimality and performance would

suggest that high performers show patterns of connectivity that are ideal for performing particular cognitive tasks.

Finally, the relationship between connectome stability and typicality may also predict cognitive performance. Individuals whose connectomes are more similar to themselves than they are to others are more easily discriminated from the group, and this uniqueness could hypothetically reflect an individualized strategy by which an individual accomplishes a task. Differences in this discriminability, defined in the current study as the ratio of one's similarity to oneself vs. one's similarity to others, may also be meaningful for predicting individual phenotypic differences. Indeed, previous work has found that connectome discriminability increases over development but is delayed in individuals with greater numbers of psychiatric symptoms (Kaufmann et al., 2017), suggesting that connectome distinctiveness may capture variance in behavior. Although complementary work has emphasized the importance of connectome discriminability in prediction such that better identifiability may improve predictive power (Amico and Goñi 2018; Elliott et al., 2019), other research has demonstrated more limited utility of connectome individuation for phenotype prediction (Finn et al., 2017; Noble et al., 2017; Greene et al., 2018) and suggested that the functional connections that contribute to identifiability are distinct from those that best predict behavior (Mantwill et al., 2021). While the relationship between discriminability and prediction is an ongoing topic of discussion, the question of whether greater discriminability relates to cognitive ability has not been explored. Here, we investigate this relationship, asking whether discriminability adds unique predictive power above and beyond its component parts, stability and typicality.

While previous literature has investigated the connectome's predictive utility, existing studies have typically focused on the strength of sets of functional connections (e.g., CPM,

machine learning approaches) or graph theoretic measures rather than connectome similarity within and between individuals. The current study tests whether overall features of the connectome— stability, typicality, optimality, and discriminability—predict cognitive performance in adults. In particular, we investigated the critical cognitive abilities of sustained attention, the ability to maintain attention over time, and working memory, the capacity to hold information in mind. We identified three datasets whose task battery included repeated fMRI runs of a sustained attention and/or working memory task. We then constructed linear models to determine which connectome features—stability, typicality, optimality, or discriminability—best predict cognitive performance across datasets. Finally, we examined whether these whole-brain features of the connectome offer additional predictive power to previously validated network-based models.

## Methods

We tested the extent to which overall features of the functional connectome, similarity to oneself (connectome stability), similarity to others (connectome typicality and optimality), and the ratio of stability to typicality (connectome discriminability, or stability ÷ typicality), were related to performance on sustained attention and working memory tasks (**Figure 1**). To evaluate the replicability of any effects, we performed analyses on three independent datasets. In each dataset, participants performed a sustained attention task during at least two fMRI runs. In two of the three datasets, participants also performed two runs of a working memory task. All data were collected with IRB approval and secondary analysis was approved by the University of Chicago Institutional Review Board.

*Data*

Dataset 1. In Dataset 1 (n = 25), described in detail in Rosenberg et al., 2016, sustained attention was measured by performance on a gradual-onset continuous performance task (gradCPT; Esterman et al., 2013). Participants (13 females, ages 18–32 years, mean age = 22.7 years) performed three gradCPT runs in the same scan session (13:44 min/run). Two 6-min resting-state runs (one before and one after the gradCPT runs) were also collected. Task data are missing or were excluded for excessive head motion from one gradCPT run for six participants. For these participants, analyses are performed on their two available runs. Additionally, one resting-state run is missing or excluded for excessive head motion for two participants.

Each gradCPT run consisted of four 3-min blocks interleaved with 32-s rest periods. During the task, grayscale images of city and mountain scenes gradually transitioned from one to the next over a period of 800 ms. Participants were instructed to respond with a button press to city scenes (90% of trials) but withhold a response to mountains (10%). Sustained attention performance was operationalized as sensitivity (d'), calculated as z(hit rate) − z(false alarm rate), averaged over each participant's two or three available runs. Behavioral inclusion criteria were defined a priori as only those participants whose performance was within 2.5 standard deviations from group mean performance. No participants were excluded from Dataset 1 based on this criterion.

MRI data were collected on a 3T Siemens Trio TIM system equipped with a 32-channel head coil. Preprocessing was performed as described in Rosenberg et al., (2016) using BioImage Suite. Data were motion corrected using SPM8. Linear and quadratic drift, mean signal from cerebrospinal fluid, white matter, and gray matter, and a 24-parameter motion model including 6

motion parameters, 6 temporal derivatives, and their squares were regressed from the data. Additionally, data were temporally smoothed with a zero mean unit variance Gaussian filter.

Dataset 2. In Dataset 2 (n = 94), volunteers (61 females, ages 18–36 years, mean age = 23.1 years) participated in two separate scanning sessions, collected ~2 weeks apart (mean = 17.2 days, SD = 20.0 days). Participants were drawn from the dataset described in Yoo et al., (2022a).

As in Dataset 1, sustained attention was assessed with performance on two runs of the gradCPT task (10 min/run). Unlike in Dataset 1, gradCPT runs were collected during different days rather than during the same scan session. Dataset 2 also included a visual short-term memory task (VSTM; two 10-min runs collected on different days), which measures visual working memory (Luck and Vogel, 1997). In the VSTM, an array of 2, 3, 4, 6, or 8 colored discs was presented on the screen for 100 ms. Following the presentation and a 900-ms fixation cross, a second array was presented either with or without a color change, with a color change occurring on 50% of trials. Participants were given 2,000 ms to respond with a button press if they detected a color change. Working memory performance on the VSTM was measured with Pashler's K (Pashler, 1988). In addition to the gradCPT and VSTM, participants performed a multiple object tracking task (MOT; 10 min/run), viewed the naturalistic movie Inscapes (Vanderwal et al., 2015, 2021; 7:16 min/run), and rested (10 min/run).

For sustained attention and working memory analyses separately, we defined inclusion criteria as only those participants whose performance was within 2.5 standard deviations from the group mean performance. No participants were excluded from Dataset 2 based on this criterion. Additionally, participants were excluded from stability and discriminability analyses if they did not complete the first or second sustained attention or working memory run because connectome stability and discriminability could not be calculated for these individuals. This resulted in a

sample size n = 65 for sustained attention and n = 72 for working memory analyses, and 56 participants were included in stability and discriminability analyses of both tasks. For analyses of connectome typicality, the full sample size n = 94 was used. Finally, connectome optimality analyses were conducted on a sample of n = 89 which corresponded to the full sample minus the top performers in either the sustained attention or working memory task, whose connectomes were averaged as the "optimal" pattern in analyses. If participants only completed one sustained attention or working memory run, task-specific functional connectivity was calculated using one run. Otherwise, functional connectivity matrices were averaged between sustained attention or working memory runs.

Functional MRI data were collected on a 3T Siemens Prisma system with a 64-channel head coil. Preprocessing was performed with AFNI (Cox 1996) and included the removal of the first three volumes and censoring of volumes with outliers in more than 10% of voxels and those for which the Euclidean norm of the head motion parameter derivatives were >0.2 mm. Data were despiked, slice-time corrected, and motion corrected. Mean signal from the CSF, white matter, and whole brain was regressed from the data, as well as 24 motion parameters. Finally, data were aligned to a high-resolution anatomical image (MPRAGE) and normalized to MNI space.

Dataset 3. Dataset 3 (n = 316) included data provided by the large-scale open-source Human Connectome Project S1200 release (HCP; Van Essen et al., 2013; Glasser et al., 2013). All fMRI data were acquired on a 3T Siemens Skyra scanner. Participants performed seven tasks (emotion, gambling, language, social, motor, N-back, and relational) and completed two resting-state runs over two-day visits. Each condition involved two runs with opposite phase encoding directions (LR and RL) which were completed in the same scanning session. Data were minimally preprocessed using HCP pipelines (Barch et al., 2013; Glasser et al., 2013). Additionally, the first

15 volumes of each run were discarded and nuisance covariates were regressed from each run, including 24 motion-related parameters (6 translational and rotational motions, 6 derivatives, and their squares), three mean tissue signals (global, white matter, and cerebrospinal fluid), and linear and quadratic trends (Yoo et al., 2022b).

Our analyses included data from 316 participants (154 females, ages 22–36+ years) who completed all nine fMRI conditions with low head motion in all (<3-mm translation, <3 degree rotation, and <0.15-mm frame-to-frame displacement), had a behavioral fluid intelligence score, and were unrelated to any of the other subjects included in the sample. Fluid intelligence scores were required for inclusion in separate analysis (Yoo et al., 2022b) but were not used in the current study. Relatedness was based on family structure verified with genetic information, and one subject was randomly selected from each family to be included in the sample.

Sustained attention performance in the HCP sample was operationalized as percent accuracy on 0-back blocks of the N-back task. During 0-back blocks, a target cue was presented at the start of the block and participants were instructed to respond with a button press any time the target reappeared. With this design, the 0-back, or target-detection, task functioned as a type of continuous performance task (in which participants make one response to stimuli in a frequent category and another to rare targets), typically used to measure sustained attention. Working memory performance was measured with performance on 2-back runs of the N-back task, during which participants responded with a button press if an image was the same as the image shown two trials previously. The N-back task consisted of two 5-min runs with eight blocks per run. Of these, four blocks were 0-back blocks and four were 2-back blocks. Both 0-back and 2-back blocks contained 10 trials per block, and a target image was displayed on two trials, with two–three

nontarget lures displayed on other trials. Images were presented for 2 s with a 500-ms intertrial interval.

As in Datasets 1 and 2, participants were excluded if behavioral performance was greater than 2.5 standard deviations below mean performance for sustained attention or working memory runs. As a result, sustained attention and working memory analyses included 304 and 311 participants, respectively, and 300 participants were included in both analyses.

*Functional connectivity calculation*

Functional connectivity was calculated in the same way for all three datasets. For each task run, functional data were parcellated using the 268-node Shen functional atlas (Shen et al., 2013) and the time series for all voxels within a parcellation were averaged. In Dataset 3, because 0-back and 2-back blocks were collected in the same fMRI run, time series were constructed by concatenating 0-back and 2-back blocks separately. A lag of 8 TRs (~5.76 s; Lee et al., 1995) was used to adjust for the hemodynamic delay. Then, the Pearson correlation was calculated between all pairs of averaged time series and Fisher z-transformed. Each value in the resulting 268 × 268 matrix, referred to as an "edge," reflects the strength of the connection between two brain regions within an individual. To test whether results depended on parcellation scheme, we also replicated our main analysis in Dataset 1 using functional connectivity matrices constructed using the 122-node Yeo functional atlas (Yeo et al., 2011). These results are included in **Appendix 1 Table 6**.

*Connectome stability*

We assessed the stability of individuals' functional connectivity patterns across fMRI runs, controlling for head motion inside the scanner. Our primary analyses focused on whether within-task stability—that is, similarity between repeated sustained attention or working memory runs—was related to performance on the respective task. As a secondary analysis, we investigated whether stability within or between unrelated task or resting-state runs reflected sustained attention and working memory performance.

To calculate connectome stability, we first vectorized connectomes by flattening the lower triangle of the symmetric 268 × 268 whole-brain functional connectivity matrix from each participant and fMRI run, resulting in a vector of 35,778 edges. Stability of sustained attention task run connectomes was calculated for each participant as the Pearson correlation between vectorized connectomes (i.e., mean pairwise correlation value for three gradCPT runs for Dataset 1, two gradCPT runs for Dataset 2, 0-back blocks from two N-back runs for Dataset 3). Stability of working memory task run connectomes was calculated for each participant as the Pearson correlation between vectorized working memory connectomes (i.e., two VSTM runs for Dataset 2, 2-back blocks from two N-back runs for Dataset 3). Additionally, overall stability values were obtained for each participant. These were calculated as the mean Pearson correlation value between all pairs of runs for a dataset. Pearson correlation coefficients were Fisher's z-transformed before averaging. Pearson correlation was used as our measure of stability in keeping with previous work (Kaufmann et al., 2017; Vanderwal et al., 2021) and because we were interested in relative rather than raw edge strength values which could vary across sessions and site due to scanner-related and other noise.

We measured the relationship between connectome stability and task performance by calculating the partial Spearman rank correlation between all pairs of runs, including terms for average frame-to-frame head displacement during both runs being correlated as well as the absolute difference in displacement between runs. As a note, we used Pearson correlation to calculate relationships between functional connectivity patterns, as we expected these values to be linearly related to one another. We used Spearman rank correlation to relate characteristics of functional connectivity patterns (e.g., stability and typicality) and to relate these characteristics with performance, as we expected these variables to be monotonically but not necessarily linearly related.

To test whether connectome stability was related to overall task performance and not consistency in task performance, we conducted an additional control analysis. Again, we calculated the partial correlation between stability and performance but included a term controlling for absolute difference in task performance between repeated runs of either the sustained attention or working memory task. As a second control, we calculated the partial Spearman rank correlation between connectome stability and performance on only one sustained attention or working memory run—i.e., the first, second, or third run, again controlling for mean head motion and change in head motion between runs. These control analyses were conducted on Datasets 1 and 2 because they required run-specific behavioral data which was unavailable for Dataset 3.

*Connectome typicality*

Do better performers show a more typical task-based connectome—i.e., a connectome that is more similar to the group average? Previous work suggests that cognitive tasks such as the N-back task elicit characteristic patterns of activation across individuals (Satterthwaite et al., 2013;

Rosenberg et al., 2020). If these patterns are reflected in the group average connectome, the extent to which an individual resembles the group may reflect task-related processing and therefore performance. Alternatively, connectome typicality may be indicative of more typical task performance such that better and worse performers have less typical patterns of connectivity. To quantify typicality, we first found an individual's mean functional connectivity pattern for a given run type by averaging connectomes across repeated runs of either sustained attention or working memory tasks. Connectomes were vectorized by extracting the lower triangle of the matrix. Finally, typicality values were calculated for each participant by computing the Pearson correlation between their vectorized connectome and the mean connectome of all other participants within each dataset. This process was repeated until every participant served as the left-out participant. This resulted in two typicality values for each participant: typicality on sustained attention runs and typicality on working memory runs. We tested whether connectome typicality was related to better or more typical task performance by calculating the correlation between connectome typicality values and participants' overall task performance scores or their absolute difference in performance from the mean.

*Connectome optimality*

Perhaps taking an average across the group adds noise and does not reflect a theoretical "optimal" task-specific pattern of connectivity. Instead, similarity to the average connectome of the top performers—rather than similarity to the group average— may be a better indicator of task performance. We tested this by calculating the similarity of each participant's mean task-specific connectome (calculated by averaging functional connectomes over repeated task runs) to the mean connectome of participants who performed best on either the sustained attention or working

memory tasks. This optimal functional connectivity pattern was calculated as the average task connectome (from either sustained attention or working memory runs) of the top 5% of performers in each task. The resulting "optimal" connectome was drawn from the top two participants in the Dataset 1 sustained attention task, and the top five performers in the Dataset 2 sustained attention and working memory tasks, respectively. In Dataset 3, 46 participants achieved the top behavioral performance score (100%) on the sustained attention task so the optimal connectome was calculated as the average across these performers. For the working memory task in Dataset 3, 19 participants achieved the top 5% of scores. Connectome optimality values (i.e., similarity of participants' mean task-specfic connectome to the optimal connectome) were not calculated for top performers included in the optimal connectome. For comparison, optimality analyses were repeated using the average of performers who achieved the single highest score and the top 10% of performers as optimality thresholds. Results from these analyses are included in **Appendix 1 Table 9**.

*Connectome discriminability*

Finally, it might be the case that performance on sustained attention and working memory tasks is best explained not by the stability of an individual's task-based connectome nor by the similarity of their connectome to others', but by a relationship between the two. For example, previous work has suggested that connectome distinctiveness, or the extent to which an individual's connectome can be identified from the group, is lower in adolescence and early adulthood for individuals with increased neurological dysfunction, such as symptoms of attention deficit disorder and depression and prodromal symptoms of schizophrenia (Kaufmann et al., 2017).

Here, we were interested in whether differences in connectome distinctiveness, which we term discriminability, are observed in the nonclinical adult populations as well and, further, whether they relate to sustained attention and working memory performance. To investigate this possibility, we calculated the discriminability of each participant's sustained attention and working memory connectomes, or the similarity of an individual's connectome across fMRI runs relative to their similarity to the group within each dataset. Discriminability was defined as the ratio of an individual's connectome stability (similarity to oneself) to their connectome typicality (similarity to all other participants) for sustained attention and working memory runs separately. Pearson r values reflecting stability and typicality were Fisher-z transformed before their ratio was taken.

Discriminability is related to subject identifiability or distinctiveness, which have previously been quantified as classification accuracy (Kaufmann et al., 2017), or whether an individual's connectivity most closely resembles oneself across runs (Finn et al., 2015). While these measures are conceptually related, each individual's connectome identifiability is a binary (because their functional connectivity pattern in one run is either most similar to their own pattern in another run or not), whereas their discriminability is continuous. In the datasets included in the current study, connectome identifiability and discriminability were significantly related for both sustained attention ($rs = 0.437$, $P = 2.14 * 10^{-20}$) and working memory runs ($rs = 0.533$, $P = 7.80 * 10^{-30}$). We use discriminability rather than identifiability here to ask whether an individual's behavior scales with their relative within-subject vs. between-subject variation in functional connectivity patterns.

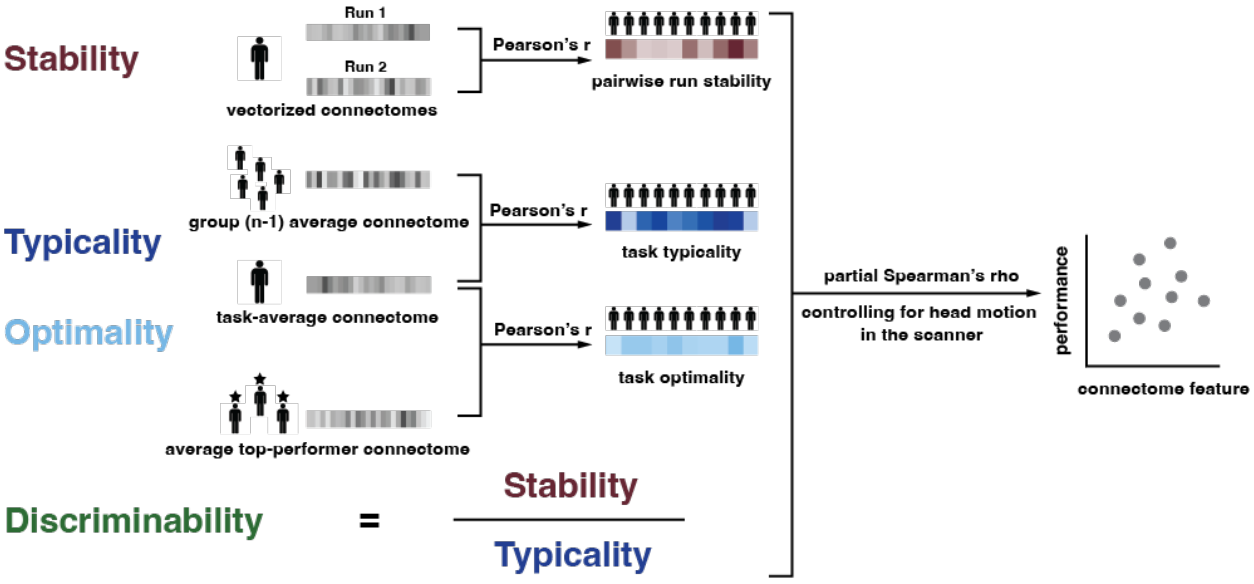



*Figure 1. Defining connectome features*. Connectome features were calculated by correlating vectorized task connectomes across runs (stability) or people (typicality and optimality). Connectome discriminability was calculated as the ratio between fisher-normalized connectome stability and typicality values. The relationship of these features with task performance was determined using partial Spearman's rank correlation accounting for in-scanner head motion.

*Modeling behavior across datasets*

We evaluated the unique variance in sustained attention and working memory task performance explained by connectome features using mixed-effects models. We constructed eight models to determine which model best predicted performance across datasets. To control for differences between datasets, behavioral performance was z-scored within dataset and stability, typicality, and discriminability values were z-scored across datasets.

We first constructed four separate models that included a predictor of stability, typicality, optimality, or discriminability with a random intercept term for dataset. Additionally, these models included covariates for mean in-scanner motion and difference in motion between runs. Next, to

examine whether similarity within and across individuals explain unique variance in performance, we constructed four additional models. To compare the unique contributions of stability and typicality, we built a model with fixed effects of stability and typicality and another model with effects of stability, typicality, and their interaction. Next, we constructed two additional models comparing stability and optimality—one model with fixed effects of stability and optimality and a second model with stability, optimality, and their interaction. Models did not include both typicality and optimality predictors because these features were highly correlated across participants (**Table 1**). Additionally, because discriminability is the ratio of stability and typicality, we did not include a discriminability term in these models. However, the interaction term of stability and typicality may be theoretically related to discriminability. Models were constructed using R's lmer function.

$$behavior = intercept + \beta 1 * (stability) + \beta 2 * (typicality) + \beta 3 * mean\ motion + \beta 4 * motion\ diff + \varepsilon$$

$$behavior = intercept + \beta 1 * (stability) + \beta 2 * (typicality) + \beta 3 * stability * typicality + \beta 3 * mean\ motion + \beta 4 * motion\ diff + \varepsilon$$

$$behavior = intercept + \beta 1 * (stability) + \beta 2 * (optimality) + \beta 3 * mean\ motion + \beta 4 * motion\ diff + \varepsilon$$

$$behavior = intercept + \beta 1 * (stability) + \beta 2 * (optimality) + \beta 3 * stability * optimality + \beta 3 * mean\ motion + \beta 4 * motion\ diff + \varepsilon$$

*Contributions of connectome features above and beyond CPMs*

In addition to testing whether connectome stability, typicality, and optimality independently predict performance on cognitive tasks, we asked whether they provide additional utility by improving the predictive power of previously established CPMs. Such models, which identify

functional connections whose strength scales with behavior and use the strength of these connections to predict behavior in novel individuals, have been validated for prediction of both sustained attention (Rosenberg et al., 2016) and working memory performance (Avery et al., 2020). To test whether connectome features of stability, typicality, and optimality improved behavioral prediction over and above that achieved by CPM network strength alone, we constructed linear models with terms for CPM network strength, connectome features, and their interaction for sustained attention and working memory runs separately.

$$behavior = intercept + \beta 1 * CPM\ network\ strength + \beta 2 * connectome\ feature + \beta 3 * CPM\ network\ strength * connectome\ feature + \varepsilon$$

CPM network strength was calculated for sustained attention runs using the sustained attention CPM (Rosenberg et al., 2016; https://github.com/monicadrosenberg/Rosenberg_PNAS2020) and for working memory runs using the working memory CPM (Avery et al., 2020). Specifically, sustained attention network strength values were computed by taking the difference between the average functional connection strength in a high-attention network (whose strength predicts better sustained attention performance) and a low-attention network (whose strength predicts worse sustained attention performance), as characterized in Rosenberg et al., (2016). Working memory strength values were calculated analogously using the predefined high- and low-working memory networks (Avery et al., 2020). Network strength and connectome feature values were z-scored within-dataset for sustained attention and working memory separately. Because the sustained attention CPM was originally trained on data from Dataset 1, the sustained attention models were tested only in Datasets 2 and 3 in the current study. Similarly, because the working memory CPM

was originally trained on data from Dataset 3, the current working memory models were tested only in Dataset 2. We compared the output of these models to models including a single term of CPM network strength or connectome feature.

*Anatomy of connectome stability and typicality*

To investigate whether the stability and typicality of functional connectivity differs throughout the brain, we calculated the stability and typicality of functional connections within and between canonical functional networks. Network parcellations were defined in Noble et al., (2017).

We next investigated the degree to which any relationship between connectome stability and task performance was driven by individual functional networks. To do so, we "computationally lesioned" networks from the connectome—that is, removed all nodes within a network from the connectivity matrix—and recalculated stability values for the lesioned connectome. We then recalculated the relationship between lesioned connectome stability and task performance iteratively for each network and evaluated the change in correlation relative to that observed with the full connectome.

We used a similar method to determine how individual functional networks contributed to the relationship between the typicality of functional connectivity patterns and task performance. Typicality was calculated after networks were computationally lesioned, and the relationship with task performance was recalculated using the lesioned connectome. The change in correlation indicates how each network contributed to the relationship observed between typicality of the full connectome and behavior.

Changes in the correlation between computationally lesioned connectome stability and task performance were compared to a distribution of 1,000 null permutations. To create the distributions, network labels were shuffled and the change in correlation between randomly lesioned connectome stability and performance was recalculated 1,000 times. Because calculating lesioned typicality values was more computationally intensive, lesioned typicality values were compared against 250 null permutations.

## Results

*Individual and group-level connectome features are related*

Connectome stability, typicality, optimality, and discriminability are not independent of one another (**Table 1**). Stability was correlated with typicality and optimality for both sustained attention and working memory runs in Datasets 2 and 3 and correlated with discriminability in all datasets. Typicality and optimality were also strongly correlated across all datasets in both sustained attention and working memory runs. Discriminability was positively correlated with typicality and optimality in Dataset 3 but negatively correlated with typicality in Dataset 2 working memory runs. Because these connectome variables are related but not perfectly correlated, they may explain unique variance in sustained attention and working memory performance.

Dataset 1

| | | Stability | | | Typicality | | | Optimality | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | rho | sig. | | rho | sig. | | rho | sig. | |
| Typicality | Sustained attention | .288 | .163 | | | | | | | |
| Optimality | Sustained attention | .492 | .017 | * | .524 | .010 | * | | | |
| Discriminability | Sustained attention | .848 | <.001 | *** | -.215 | .302 | | .245 | .260 | |

Dataset 2

| | | Stability | | | Typicality | | | Optimality | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | rho | sig. | | rho | sig. | | rho | sig. | |
| Typicality | Sustained attention | .271 | .029 | * | | | | | | |
| | Working memory | .258 | .029 | * | | | | | | |
| Optimality | Sustained attention | .483 | <.001 | *** | .854 | <.001 | *** | | | |
| | Working memory | .301 | .013 | * | .913 | <.001 | *** | | | |
| Discriminability | Sustained attention | .831 | <.001 | *** | -.229 | .067 | | .092 | .475 | |
| | Working memory | .808 | <.001 | *** | -.266 | .024 | * | -.144 | .240 | |

Dataset 3

| | | Stability | | | Typicality | | | Optimality | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | rho | sig. | | rho | sig. | | rho | sig. | |
| Typicality | Sustained attention | .708 | <.001 | *** | | | | | | |
| | Working memory | .677 | <.001 | *** | | | | | | |
| Optimality | Sustained attention | .696 | <.001 | *** | .984 | <.001 | *** | | | |
| | Working memory | .665 | <.001 | *** | .985 | <.001 | *** | | | |
| Discriminability | Sustained attention | .874 | <.001 | *** | .356 | <.001 | *** | .337 | <.001 | *** |
| | Working memory | .916 | <.001 | *** | .404 | <.001 | *** | .403 | <.001 | *** |

*Table 1. Across-subject Spearman's correlation values between connectome features.* Significance values are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

*Within-task connectome stability uniquely predicts behavior*

We first investigated whether stability of an individual's functional connectome during sustained attention and working memory tasks predicts task performance in three independent

datasets. To do so, we calculated the partial correlation between connectome stability and task performance controlling for mean head motion on both runs as well as the difference in mean head motion between runs (**Figure 2**).

Mean connectome stability and sustained attention performance were positively correlated in all three datasets, and significant in Datasets 2 and 3 (DS1: partial rs = 0.344, P = 0.109; DS2: partial rs = 0.411, P = 8.22∗10−4; DS3: partial rs = 0.168; P = 3.34∗10−3; **Figure 2**; **Appendix 1 Table 7**). For Datasets 2 and 3, which also included working memory tasks, connectome stability during working memory runs was significantly related to working memory performance (DS2: partial rs = 0.264, P = 0.027; DS3: partial rs = 0.207; P = 2.54∗10−4; **Figure 2**; **Appendix 1 Table 7**). These results suggest that individuals who express more similar functional connectivity patterns across repeated fMRI runs of a sustained attention or working memory task tend to perform better on these tasks. Additionally, connectome stability during sustained attention tasks was significantly correlated with stability during working memory tasks in both datasets that included these tasks after controlling for mean head motion and difference in head motion between runs (DS2: partial rs = 0.337, P = 0.013; DS3 partial rs = 0.693, P = 3.28∗10−46), suggesting that participants tend to have similar levels of connectome stability between sustained attention and working memory tasks. Note, however, that in Dataset 3, sustained attention and working memory tasks were performed in interleaved blocks, potentially inflating estimates of connectome stability similarity across task types.

Perhaps higher performers have more stable functional connectivity patterns regardless of task. We next tested whether it was stability within sustained attention and working memory runs specifically or across-run stability more generally that predicted performance. To do so, we calculated stability between all pairs of fMRI runs within a dataset and again found the partial

correlation between stability and performance controlling for head motion (**Figure 3**). To control for the increased likelihood of observing positive but spurious correlations when performing multiple comparisons, we note here which correlations survive Bonferroni correction.

Mean stability during resting-state runs was not significantly related to sustained attention performance (DS1: partial rs = 0.260, P = 0.269; DS2: partial rs = 0.150, P = 0.168; DS3: partial rs = 3.80∗10−3, P = 0.948) nor working memory performance (DS2: partial rs = −.088, P = 0.422; DS3: partial rs = 0.076, P = 0.185; **Figure 3**). This suggests that the relationship between connectome stability and performance is relatively specific to task-based functional connectivity.

Stability between resting-state and task runs was not reliably related to either sustained attention or working memory performance across datasets. Average stability in functional connectivity between rest and sustained attention runs (i.e., less change from rest to task runs) predicted sustained attention performance in Dataset 3 only (DS1: rs = 5.64∗10−4, P = 0.998; DS2: rs = 0.137, P = 0.197; DS3: rs = 0.115, P = 0.046). Similarly, smaller changes between rest and working memory runs predicted working memory performance in Dataset 3 but not Dataset 2 (DS2: rs = −.029., P = 0.787; DS3: rs = 0.118, P = 0.037). While previous work using data from the Human Connectome Project found that similarity between resting-state and task FC is related to task performance (Schultz and Cole, 2016), we only observed this relationship in Dataset 3, a subsample of the Human Connectome Project dataset. We did not observe this relationship consistently across other datasets.

In some cases, stability during different tasks was predictive of performance on either the sustained attention or working memory tasks. Specifically, in Dataset 2, stability during the MOT task, which taxes attention, predicted both sustained attention and working memory performance (Sustained attention: partial rs = 0.297, P = 0.015; Working memory: partial rs = 0.350, P =

3.67∗10−3; **Figure 3**). However, these relationships do not survive Bonferroni correction for 45 comparisons ((Ntasks2 − Ntasks)/2). In Dataset 3, connectome stability during 2- back, motor, language, and social cognition tasks was related to sustained attention performance (2-back: partial rs = 0.168, P = 3.37∗10−3; motor: partial rs = 0.116, P = 0.044; language: partial rs = 0.147, P = 0.011; social: partial rs = 0.141, P = 0.015). Connectome stability during 0-back, motor, language, and social cognition tasks similarly predicted working memory performance in Dataset 3 (0-back: partial rs = 0.219, P = 1.03∗10−4; motor: partial rs = 0.236, P = 2.82∗10−5; language: partial rs = 0.271, P = 1.38∗10−6; social: partial rs = 0.227, P = 5.75∗10−5). Only correlations with working memory performance (0-back, language, social, and motor tasks) survive Bonferroni correction. Taken together, these results suggest that the relationship between performance and stability is specific to stability in certain tasks.

Finally, to examine whether performance was related to stability more broadly—that is, across all run types—we correlated performance with a more general measure of stability for each participant. To evaluate participants' overall connectome stability, we calculated the mean stability between all pairs of runs in each dataset. We then calculated the partial Spearman rank correlation between stability and performance, controlling for individuals' mean head motion on all runs. Overall stability was not related to sustained attention performance in Dataset 1 or 2 but was marginally related in Dataset 3 (DS1: partial rs = 0.205, P = 0.336; DS2: partial rs = 0.167, P = 0.109; DS3: partial rs = 0.106, P = 0.065). For working memory performance, there was no relationship observed in Dataset 2, but there was a significant correlation with overall stability in Dataset 3 (DS2: partial rs = 0.045, P = 0.670; DS3: partial rs = 0.160, P = 4.70∗10−3). While inconsistent, the strength of the relationship between performance and overall stability increased with the total number of runs included in the calculation of overall stability. Future work can

investigate whether individuals who perform better on sustained attention and working memory tasks express a more stable pattern of functional connectivity across a variety of cognitive and task states.

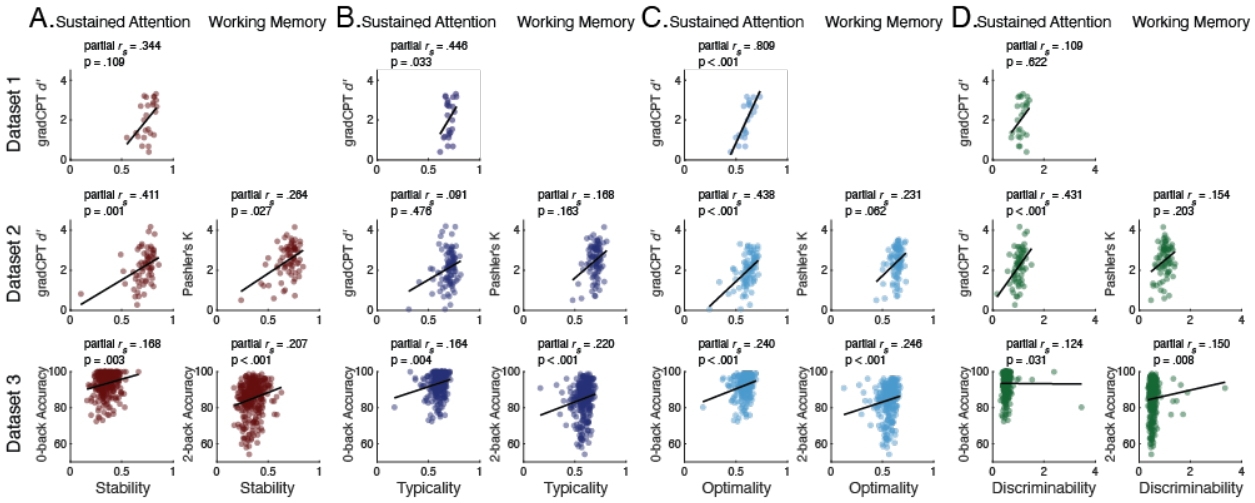



*Figure 2. Connectome features are related to sustained attention and working memory.* A) Connectome stability, B) typicality, C) optimality, D) and discriminability were differentially related sustained attention and working memory performance across datasets. Spearman's rank correlations were used to mitigate effects of outliers on the relationship between connectome features and performance.

*Connectome stability predicts behavior beyond task performance consistency*

An alternative explanation of these results is that instead of reflecting overall performance, connectome stability is related to consistency in performance, which in turn relates to overall performance. We performed two control analyses to test this account. These analyses were not performed in Dataset 3 because this sample lacked run-specific behavioral measures.

First, we recalculated the partial Spearman rank correlation between connectome stability and performance controlling for the absolute difference in performance between runs. Absolute difference in performance was inversely related to sustained attention (DS1: r = −0.301, P = 0.144; DS2: r = −0.051; P = 0.688) and working memory performance (DS2: r = −0.401, P = 4.78∗10−4) indicating that better performers were indeed more consistent (i.e., had less change in performance) between runs. After controlling for consistency in sustained attention performance, the correlation between connectome stability and performance remains positive in Dataset 1 and significant in Dataset 2 (DS1: partial rs = 0.325, P = 0.140; DS2: partial rs = 0.445, P = 2.96∗10−4). After controlling for consistency in working memory performance, the correlation between connectome stability and performance numerically weakens in Dataset 2 (partial rs = 0.198, P = 0.103).

Second, we calculated the partial Spearman rank correlation between connectome stability and performance on a single run from each task. The relationship between connectome stability and sustained attention task performance was positive for each of the three runs in Dataset 1 (run 1 partial rs = 0.431, P = 0.051; run 2 partial rs = 0.467, P = 0.028; run 3 partial rs = 0.120, P = 0.604) and for both runs in Dataset 2 (run 1 partial rs = 0.231, P = 0.068; run 2 partial rs = 0.522, P = 1.16∗10−5). For working memory runs in Dataset 2, the partial correlation between connectome stability and individual run performance was also positive, although numerically weaker than the relationship with average working memory performance (run 1 partial rs = 0.226, P = 0.060; run 2 partial rs = 0.179, P = 0.139). In combination, these control analyses suggest that relationships between taskbased connectome stability and sustained attention and working memory task performance may be partially, but not fully, driven by the fact that better performers tend to be more consistent performers.

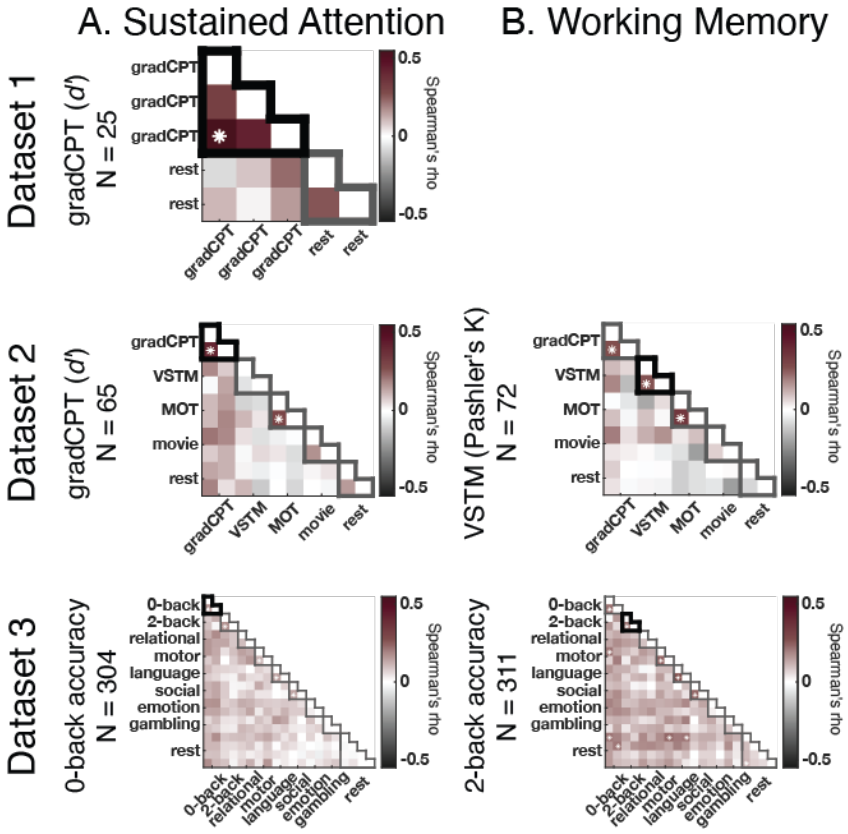



*Figure 3. Connectome stability predictions are task-specific.* Stability between pairs of runs differentially predicted a) sustained attention and b) working memory performance. Colors indicate the partial Spearman correlation between pairwise run stability and performance, controlling for head motion within and between runs. Repeated task runs are outlined in gray. Repeated task runs whose performance was used as the predicted measure are indicated with heavy black lines. Stars within black and gray outlines represent uncorrected significant correlation values, while all stars outside of the black and gray outlines represent Bonferroni-corrected significance values.

*Connectome typicality also predicts task performance*

It is possible that individuals express patterns of functional connectivity specific to particular cognitive tasks. If this is the case, we may expect a group-averaged task-based connectome to partially reflect a task-engaged pattern and the extent to which a participant's connectome resembles the group average to reflect the degree to which they achieve this characteristic task pattern. Thus, it may be that behavioral performance is related to connectome typicality, or how similar an individual is to the rest of the group.

To test this possibility, we calculated the typicality of functional connectivity patterns for each dataset and computed the partial correlation between connectome typicality and task performance, again controlling for head motion in the scanner. For sustained attention runs, connectome typicality was significantly related to performance in Datasets 1 and 3. In Dataset 2, however, there was no significant relationship between connectome typicality and sustained attention performance (DS1: partial rs = 0.446, P = 0.033; DS2: partial rs = 0.091, P = 0.476; DS3: partial rs = 0.164, P = 3.34∗10−3; **Figure 2**; **Appendix 1 Table 7**). In working memory runs, connectome typicality was not related to working memory performance in Dataset 2 but was significantly related to performance in Dataset 3 (DS2: partial rs = 0.168, P = 0.163; DS3: partial rs = 0.220, P = 2.54∗10−4). Similar to analyses of connectome stability, the strongest positive relationships between connectome typicality and task performance are largely specific to typicality during sustained attention and working memory tasks (**Appendix 1 Figure 36**), supporting the hypothesis that some task-specific patterns of behavior are captured in the group average connectome.

The inconsistency of the relationship between connectome typicality and performance, however, suggests that better performers may not simply be expressing a pattern of connectivity

resembling the group average. Alternatively, it may be the case that a more typical pattern of connectivity is reflective of more typical performance, in which case better performers would deviate more from the group average. To further investigate these hypotheses of how functional connectivity typicality relates to performance, we performed two additional analyses.

First, we tested the possibility that connectome typicality does not relate to better performance but instead relates to more typical performance. For this analysis, we measured the relationship between participants' average task-specific connectome and their absolute difference from the mean performance score. We found little evidence for this relationship across datasets for both sustained attention (DS1: partial rs = −0.153, P = 0.497; DS2: partial rs = −0.228, P = 0.073; DS3: partial rs = −0.001, P = 0.984) and working memory tasks (DS2: partial rs = −0.062, P = 0.610; DS3: partial rs = 0.071, P = 0.210), suggesting that more typical connectomes are not reliable markers of more typical performance. Furthermore, average connectome typicality values generally increase across task performance quartiles (**Appendix 1 Figure 37**), indicating that better performers are more similar to the group average than low performers.

*Similarity to the optimal performer predicts behavior*

Another explanation for the inconsistent relationship between connectome typicality and performance may be that the group average connectome is not the best proxy for task-specific patterns of functional connectivity. Instead, similarity to the functional connectivity patterns of the best performers may better reflect adherence to a task-specific connectivity pattern. To test this possibility, we examined whether similarity to the connectomes of the best performers was predictive of performance. To do so, we calculated the similarity of each participant's average task functional connectivity pattern to the mean connectivity pattern of the participants who achieved

the top 5% of scores. Then, we calculated the Spearman's rank correlation between this similarity to the "optimal" connectome and task performance. Across all datasets with sustained attention runs, similarity to the optimal connectome predicted performance during sustained attention runs (DS1: partial rs = 0.809, P = 9.10∗10−6; DS2: partial rs = 0.438, P = 4.60∗10−4; DS3: rs = 0.240, P = 1.06∗10−4; **Figure 2**; **Appendix 1 Table 7**). For working memory as well, similarity to the optimal pattern of connectivity was positively correlated with task performance in both Dataset 2 and 3, and significant in Dataset 3 (DS2: partial rs = 0.231, P = 0.062; DS3: rs = 0.246, P = 2.35∗10−5). This consistency across datasets and tasks suggests that some patterns of connectivity may be "optimal" for specific tasks, such that the extent to which an individual's connectome resembles this optimal pattern reflects performance on that task.

Do "optimal" patterns of connectivity generalize across task and dataset? We tested whether similarity to the top performer connectome of another dataset similarity predicted task performance within sustained attention and working memory domains. We observed limited generalization such that similarity to the top performers in Dataset 2 significantly predicted gradCPT performance in Dataset 1 (partial rs = 0.542, P = 7.55∗10−3) but not 0-back performance in Dataset 3 (partial rs = 0.098, P = 0.088). Similarity to the mean optimal connectome in Dataset 1 predicted 0-back performance in Dataset 3 (partial rs = 0.183, P = 1.37∗10−3) and, conversely, similarity to the optimal 0-back connectome in Dataset 3 predicted gradCPT performance in Dataset 1 (partial rs = 0.455, P = 0.029). However, gradCPT performance in Dataset 2 was not predicted by similarity to the top connectomes of either Dataset 1 (partial rs = −0.004, P = 0.972) or Dataset 3 (partial rs = −0.093, P = 0.470). For working memory tasks, similarity to the top performers' connectome in Dataset 2 did not generalize to predict 2-back performance in Dataset 3 (partial rs = 0.078, P = 0.171) nor did similarity to the top 2-back connectome predict VSTM

performance in Dataset 2 (partial rs = 0.177, P = 0.143). Full comparisons are included in **Appendix 1 Table 8**. Thus, "optimal" functional connectivity patterns may be relatively task and/or dataset-specific.

*Connectome discriminability inconsistently predicts behavior*

Previous work demonstrated increases in connectome distinctiveness across development and delays in this trajectory in clinical populations (Kaufmann et al., 2017). These findings raise the possibility that individuals who perform better on cognitive and attentional tasks are achieving a more unique connectivity pattern, such that their connectomes are more similar to themselves than they are to the group. To investigate this, we calculated each individual's discriminability, or the ratio of their connectome stability to typicality on either sustained attention or working memory runs. While discriminability values could theoretically range from negative infinity to positive infinity, in the current study, values ranged from 0.173 to 3.46. For sustained attention runs, connectome discriminability was not related to performance in Dataset 1 but did significantly predict performance in Datasets 2 and 3 (DS1: partial rs = 0.109, P = 0.622; DS2: partial rs = 0.431, P = 4.17∗10−4; DS3: partial rs = 0.124; P = 0.031; **Figure 2**; **Appendix 1 Table 7**). In working memory runs, connectome discriminability did not predict performance in Dataset 2 but was significantly related to working memory performance in Dataset 3 (DS2: partial rs = 0.154, P = 0.203; DS3: partial rs = 0.150; P = 8.33∗10−3). Although there is some evidence for a relationship between connectome discriminability and performance, inconsistency across datasets suggests that better performers may not simply be expressing more unique—i.e., identifiable—patterns of connectivity.

*Models based on connectome stability and optimality best predict behavior*

We constructed four linear models, each containing one feature of the connectome (i.e., stability, typicality, optimality, and discriminability) to evaluate how each feature explained variation in sustained attention and working memory performance separately. Model output is summarized in **Table 2**. To determine which of our candidate models best explained variations in both sustained attention and working memory performance, we compared the Akaike Information Criteria (AIC) values of each model. AIC values quantify to what extent a model minimizes both bias and variance, such that lower AIC values indicate a more optimal minimization (Burnham and Anderson, 2004). AIC values are unitless but enable the comparison of models fitting the same data to determine which candidate model least overfits the data. Using these criteria for models explaining sustained attention performance, we observed the lowest AIC value for the model containing only a fixed effect for optimality. A model containing only a stability or typicality term also significantly captured variability in sustained attention performance. When predicting working memory performance, we observed the lowest AIC value for the stability model, suggesting that this is the bestfit model for predicting working memory performance. However, typicality, optimality, and discriminability models explained significant variability in working memory performance.

We were further interested in whether interactions between variables explained additional variance in performance. To determine whether similarity within and between individuals capture unique variability in performance, we constructed four additional models. Two models included terms for stability and typicality and stability, typicality, and their interaction. An additional two models compared stability and optimality, and stability, optimality, plus their interaction. Importantly, discriminability was not included in these models because it was derived from

stability and typicality variables and typicality and optimality were not include in the same model to avoid multicollinearity. Results indicated that, for sustained attention runs, a model with fixed effects for stability and optimality best predicted sustained attention performance of these additional models. However, only the fixed effect of optimality was a significant predictor in this model. Similarly, for working memory runs, a model with effects of stability and optimality best predicted working memory performance of these models. However, only stability was a significant predictor in this model. These results suggest that a model considering connectome similarity within (stability) and across (optimality) participants may be the most useful predictor of cognitive performance.

| | Sustained Attention | | | | | Working Memory | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | coefficient | std error | significance | | AIC | coefficient | std error | significance | | AIC |
| Stability | .4384 | .1055 | $1.00*10^{-4}$ | *** | 951.0 | .4054 | .0913 | $1.59*10^{-5}$ | *** | **1005.2** |
| Typicality | .2157 | .0589 | $5.22*10^{-3}$ | ** | 952.4 | .1801 | .0611 | $5.72*10^{-3}$ | ** | 1014.4 |
| Optimality | .3205 | .0524 | $2.63*10^{-9}$ | *** | **930.3** | .1728 | .0548 | $3.27*10^{-3}$ | ** | 1012.7 |
| Discriminability | .0807 | .0514 | .1170 | | 963.5 | .1088 | .0519 | .0969 | | 1017.7 |
| Stability | .3175 | .1206 | .0101 | * | 952.4 | .3498 | .0986 | $4.59*10^{-4}$ | *** | 1008.1 |
| Typicality | .1548 | .0793 | .0519 | | | .1105 | .0693 | .1118 | | |
| Stability | .3391 | .1230 | $6.96*10^{-3}$ | ** | 957.7 | .3888 | .1033 | $2.09*10^{-4}$ | *** | 1012.8 |
| Typicality | .1583 | .0796 | .0476 | * | | .0977 | .0704 | .1659 | | |
| Stability * Typicality | .0411 | .0627 | .5127 | | | .0650 | .0579 | .2629 | | |
| Stability | .0997 | .0922 | .3549 | | 935.6 | .3430 | .0967 | $4.81*10^{-4}$ | *** | 1007.1 |
| Optimality | .3043 | .0625 | $2.73*10^{-6}$ | *** | | .1202 | .0614 | .0510 | | |
| Stability | .1973 | .1131 | .1010 | | 938.3 | .3746 | .1031 | $3.55*10^{-4}$ | *** | 1012.4 |
| Optimality | .2833 | .0641 | $1.54*10^{-5}$ | *** | | .1086 | .0632 | .0865 | | |
| Stability * Optimality | .1007 | .0506 | .0476 | * | | .0451 | .0575 | .4335 | | |

*Table 2. Output from mixed-effects models including model weights (coefficient) and residual standard error (std error) for connectome feature predictors.* Bolded AIC values indicate the best model for predicting sustained attention and working memory performance. Significance levels are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

*Connectome features are generalizable predictors*

Do connectome features capture unique variance in behavioral prediction when used in combination with previously validated network models of sustained attention and working memory? We constructed linear models with terms for CPM network strength, stability, typicality, and their interactions to predict sustained attention and working memory performance separately. Average and mean difference in in-scanner motion between runs were also included as covariates in all models. CPM network strength was calculated using the previously published sustained attention CPM (Rosenberg et al., 2016) for sustained attention runs and previously published working memory CPM (Avery et al., 2020) for working memory runs. Model results are summarized in **Appendix 1** (**Appendix 1 Table 10**).

CPM network strength alone was the best predictor of sustained attention performance, as measured by the gradCPT in Dataset 2. Models with a single term of connectome stability or optimality predicted gradCPT performance in this dataset but were less optimal models based on AIC values. In Dataset 3, connectome feature models predicted 0-back performance better than sustained attention CPM strength. Sustained attention CPM strength alone did not significantly predict 0-back performance in the present study. However, recently published work found that sustained attention CPM strength did significantly predict 0-back performance in a larger HCP sample (Kardan et al., 2022). Differences in how data were preprocessed as well as fewer participants (n = 316 vs. n = 754) and higher and less variable behavioral scores in the current sample may have contributed to differences in model prediction between datasets. Instead, the best model for predicting 0-back performance in Dataset 3 was the model that included terms for sustained attention CPM strength, optimality, and their interactions. However, the only terms that

significantly predicted performance in this model were connectome optimality and the interaction between CPM strength and optimality.

In working memory runs in Dataset 2, a model including working memory CPM strength, connectome stability, and their interaction was the best predictor of VSTM working memory performance. However, only the connectome stability and interaction terms had significant beta coefficients in this model. In combination, these results suggest that connectome features of stability and optimality may be reliable predictors of sustained attention and working memory performance in datasets where supervised CPM models do not generalize. Furthermore, in datasets where pretrained models do not generalize on their own, incorporating connectome stability or optimality as model predictors may improve model performance.

*Individual network contributions to behavioral prediction vary*

We were interested in exploring whether the relationship between connectome stability and cognitive performance relies on whole-brain stability, or whether certain networks contribute to this relationship more than others. First, we tested whether stability in CPM networks, a subset of the connectome edges whose strength predicts sustained attention or working memory performance, also predicts performance. Stability in the sustained attention CPM network was not significantly related to sustained attention performance in any of our three datasets (DS1: rs = 0.256, P = 0.238; DS2: rs = 0.083, P = 0.518; DS3: rs = 0.058, P = 0.315). Working memory CPM network stability predicted 2-back performance in Dataset 3 but not VSTM performance in Dataset 2 (DS2: rs = 0.218, P = 0.070; DS3: rs = 0.383, P = 2.94∗10−12). These results suggest that stability of the whole-brain connectome may be more predictive than stability in CPM networks.

Are certain functional networks more stable or typical than others? We visualized within- and between-network connections in terms of their average stability and typicality (**Figure 4**) across individuals within a dataset. Stability and typicality values were normalized based on network size. This visualization revealed high consistency in both network stability and typicality values for both sustained attention and working memory task connectivity patterns. In particular, visual networks were both highly stable and highly typical in both tasks across datasets. Matrices visualizing within- and between-network connections for additional connectome features optimality and discriminability are included in **Appendix 1** (**Appendix 1 Fig 3**).

Next, we computationally lesioned all nodes belonging to each of 10 canonical networks and recalculated the correlation between stability in the remaining nodes and either sustained attention or working memory performance. The change between intact and lesioned correlation values is plotted in **Figure 5**.

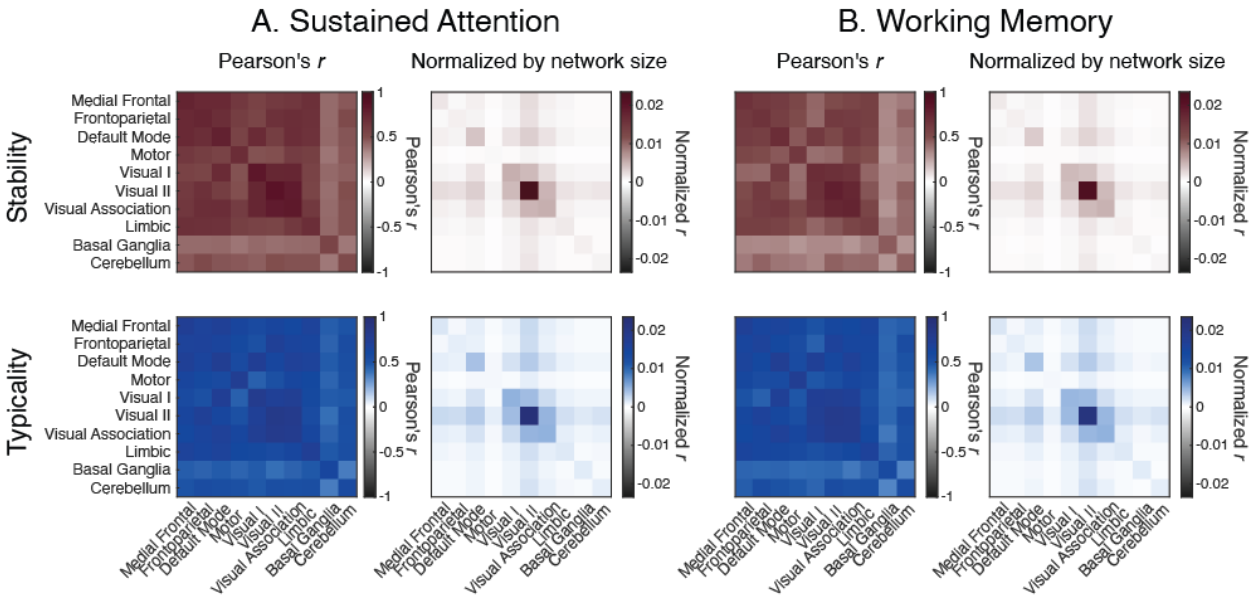



*Figure 4. Within-and between network connectome stability and typicality*. Stability and typicality of within- and between-network connections for A) sustained attention and B) working memory task connectomes. Matrices on the left represent average Pearson's correlations across datasets, while matrices on the right are average correlations normalized by network size.

Networks inconsistently contributed to the relationship between stability and cognitive performance. No network significantly contributed to the correlation between stability and sustained attention performance across all datasets. The visual I network significantly added noise to this correlation in Datasets 1 and 2 which both measured sustained attention performance using the gradCPT task. The visual association network significantly contributed to the correlation between stability and sustained attention performance in Datasets 1 and 2 but displayed a contradictory pattern of results for working memory, contributing to the relationship between stability and working memory performance in Dataset 2 but adding significant noise in Dataset 3. These results suggest that the relationship between connectome stability and sustained attention and working memory performance does not rely on specific contributions from individual

networks. Instead, networks appear to contribute in a pattern that is more task specific rather than cognitive process-specific.

We were similarly interested in whether the relationship between connectome typicality and cognitive performance relies on particular networks. We again computationally lesioned 10 canonical networks and recalculated the relationship between typicality in the intact networks and either sustained attention or working memory performance for each dataset. The change in correlation between intact and lesioned connectome typicality and cognitive performance is shown in **Figure 6**.

Network contribution to the correlation between typicality and cognitive performance largely differed across datasets. However, we observed significant contribution of the frontoparietal network to the relationship between connectome typicality and performance in working memory runs in both Dataset 2 and Dataset 3. This result is in accordance with previous work that found frontoparietal activity to be a biomarker of working memory performance in preadolescents (Rosenberg et al., 2020). The medial frontal network significantly contributed to the relationship between connectome typicality and sustained attention performance in Datasets 1 and 2 but interestingly added significant noise to this relationship in Dataset 3. Again, it is possible that network contribution to the relationship between connectome typicality and cognitive performance is specific to the task being used to measure a given cognitive ability.

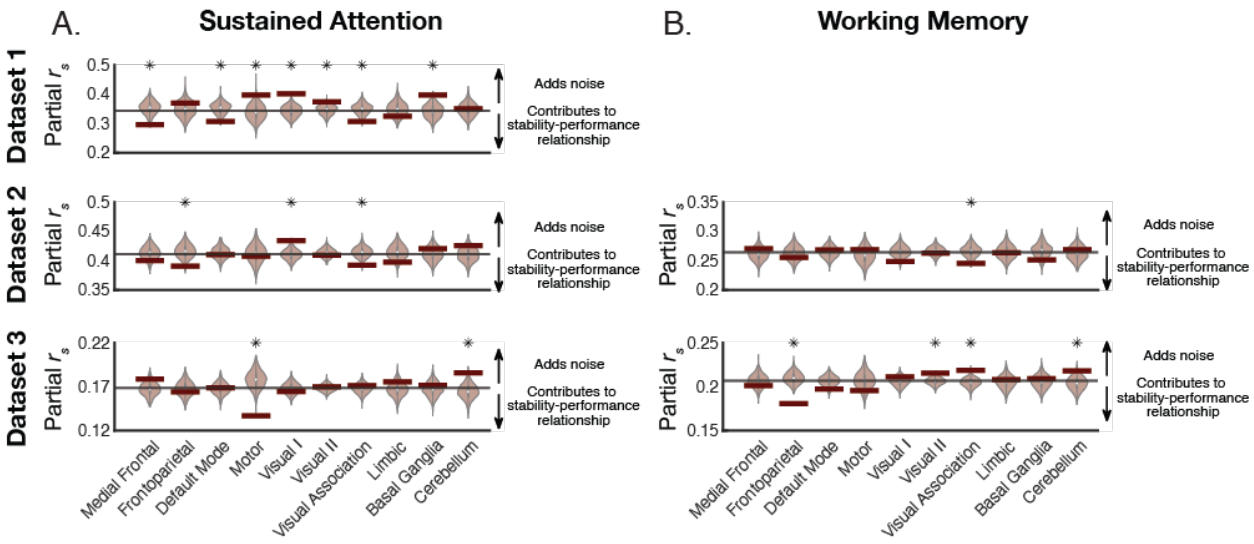



*Figure 5. Canonical networks differentially contribute to the correlation between connectome stability and cognitive performance.* Dashed lines indicate correlation values between intact connectome stability and performance. Heavy red lines indicate the Spearman rank correlation between lesioned connectomes and A) sustained attention and B) working memory performance. Decreased lesioned correlations indicate that the network contributed to the relationship between connectome stability and performance, while increased lesioned correlations suggest that the network added noise. Significance $P < 0.05$ (uncorrected) is indicated with asterisks

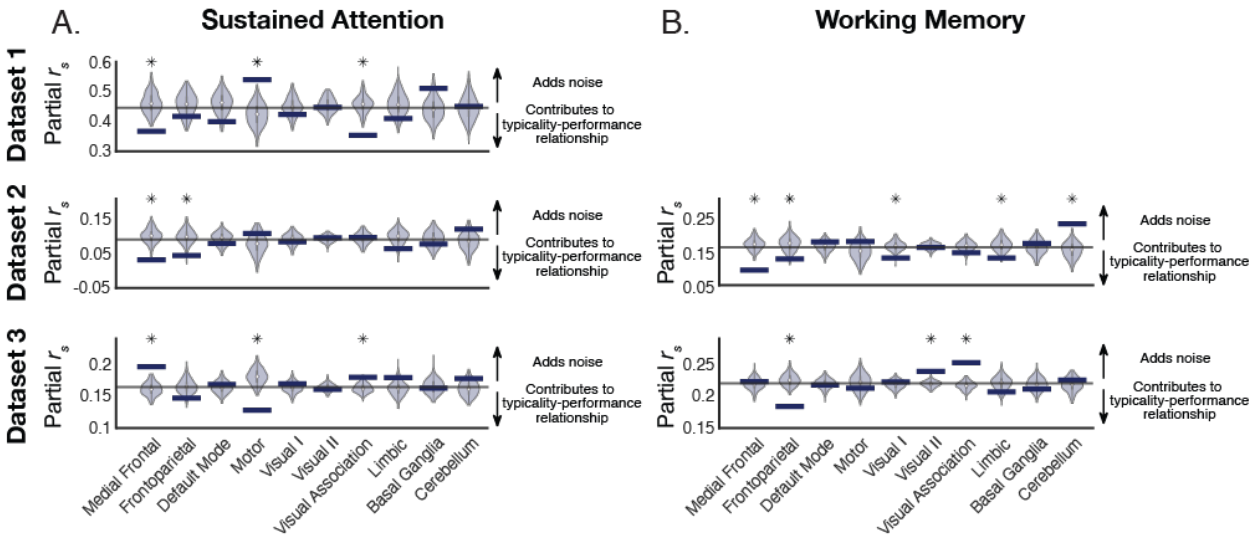



*Figure 6. Canonical networks differentially contribute to the correlation between connectome typicality and cognitive performance.* Dashed lines indicate correlation values between intact connectome typicality and performance. Heavy blue lines indicate the Spearman rank correlation between lesioned connectomes and A) sustained attention and B) working memory performance. Decreased lesioned correlations indicate that the network contributed to the relationship between connectome typicality and performance, while increased lesioned correlations suggest that the network added noise. Significance $P < 0.05$ (uncorrected) is indicated with asterisks.

## Discussion

The current study investigated whether features of the functional connectome predict individual phenotypes. Connectome stability, or similarity in functional connectivity patterns across runs, may reflect a more consistent on-task state and has been related to phenotypes in development (Kaufmann et al., 2017; Vanderwal et al., 2021; Fu et al., 2022). More typical connectomes may reflect a characteristic task-specific functional connectivity pattern that is captured in the group-average. Alternatively, this optimal pattern of connectivity may be best

captured in the mean connectome of the highest task performers. Finally, the ratio of stability to typicality (discriminability) reflects a connectome's uniqueness and has been related to clinical symptoms in development (Kaufmann et al., 2017). We compared the predictive ability of these features for sustained attention across three independent datasets and working memory across two datasets. Our findings suggest that connectome features, particularly connectome stability and optimality, can serve as generalizable predictors of sustained attention and working memory abilities.

Connectome stability was a consistent predictor of both sustained attention and working memory performance. This relationship was relatively specific to within-task (i.e., sustained attention or working memory) stability, such that the strongest relationships between connectome stability and task were observed when stability was calculated between repeated sustained attention or working memory runs. This relationship was observed across all datasets in the current study, with stability during attention tasks positively correlated with attentional performance and stability during working memory tasks positively correlated with working memory performance. Better performers in both tasks expressed more similar functional connectivity patterns across scans, even when scans were separated by weeks as in Dataset 2. Stability during different tasks (e.g., an MOT task in Dataset 2; language, social, and motor tasks in Dataset 3) was also at times related to task performance. We hypothesize that connectome stability reflects the extent to which an individual's cognitive state changes over time, with greater stability indicating a more similar cognitive state between fMRI runs. Therefore, connectome stability may indicate a more on-task state which may also be reflected in task performance.

Our feature of connectome stability is related to test–retest reliability of the functional connectome across scan sessions (Noble et al., 2019). However, while test–retest reliability tests

the extent to which functional connectivity as a measure remains stable over time, here, we aim to leverage meaningful differences in test–retest reliability across individuals to tell us something about cognitive ability. Future studies of reliability may consider to what extent a lack of test–retest reliability reflects methodological noise versus meaningful variability across tasks and individuals.

Notably, we found a lack of predictive ability from connectome stability during rest and between task and rest runs. The latter observation is in contrast to previous work that found that more similar rest and task connectivity patterns was related to higher cognition, including better working memory performance (Schultz and Cole, 2016). The current results also demonstrate a benefit of utilizing task-based functional connectivity data as a predictive tool, rather than resting-state which has been used widely in the field. For the current study, we consider what features of the functional connectomes themselves contribute to tasks' predictive ability. One possibility is that the variability of connectivity patterns during specific tasks boosts within-individual similarity's predictive power. However, a visualization of connectome standard deviation across participants suggests that tasks whose stability predicted performance and those that do not are similarly variable (**Appendix 1 Figure 39**). Another possibility is that predictive tasks induce a more variable functional connectivity pattern within-individuals than other tasks, thereby allowing stability to be a more reliable measure. However, a comparison of mean task functional connectivity standard deviation does not support this hypothesis (**Appendix 1 Figure 40**). Finally, scan length may affect the overall stability values, which may in turn affect behavior prediction. However, while stability between task runs was higher for longer tasks across datasets (correlation between number of TRs and task stability: $r = 0.675$, $P = 4.04 * 10-4$), overall task stability did not reliably lead to better prediction of sustained attention (relationship between task stability and

stability-behavior correlation: r = 0.236, P = 0.278) nor working memory (r = −0.304, P = 0.158) performance (**Appendix 1 Table 11**). Therefore, while longer scans may lead to more stable patterns of connectivity, more stable run types are not always most predictive of behavior. Future work may seek to further investigate task-based connectivity's predictive advantage. However, the current results add to the growing literature emphasizing the benefit of task-specific over resting state functional connectivity as a predictive tool.

We found initial—albeit less consistent—evidence that connectome typicality, or the extent to which an individual resembles the rest of the group, predicts cognitive and attentional performance. Tasks have been shown to induce characteristic changes in the connectome (Shine et al., 2016; Lynch et al., 2018; Greene et al., 2020), and children and adults who show activity in canonical task-positive regions, such as regions of the frontoparietal network, perform better on attention and working memory tasks (Rypma et al., 2002; Satterthwaite et al., 2013; Rosenberg et al., 2020). Thus, we hypothesized that if this task-relevant connectivity is reflected in the group-average connectome, the extent to which an individual resembles the group may scale with their task engagement and performance. In the present datasets, connectome typicality was related to sustained attention performance in two of three datasets and working memory performance in one of two datasets. A linear model with a single fixed effect of typicality significantly predicted both sustained attention and working memory across datasets. These results suggests that individuals who express shared, task-specific connectivity patterns may indeed show more on-task cognitive performance as evidenced by higher behavioral scores.

Following these initial inconsistent results, we further explored how connectome typicality might relate to behavior. We did not find evidence that connectome typicality is related to typicality in performance. Instead, we found that, for both sustained attention and working memory

tasks, connectome optimality—or, similarity to the connectome of the highest performer on a given task—was positively correlated with task performance across all datasets. In combination with the previous analysis, these results support the notion of optimal task-specific connectivity patterns but suggest that these patterns may become obscured when averaged across participants. We also observed limited generalizability of connectome optimality, such that the similarity to the optimal sustained attention or working memory connectome in one dataset did not reliably predict sustained attention or working memory performance in other datasets, respectively. Future studies may seek to further characterize these "optimal" patterns of connectivity and explore the extent of their generalizability across contexts and tasks.

Finally, we tested the predictive utility of connectome discriminability, or an individual's ability to be distinguished from others. To do so, we defined connectome discriminability as the extent to which an individual looks more like themselves than others or, intuitively, the ratio of connectome stability and typicality. This measure is related but not identical to previous work that has characterized connectome identifiability using fingerprinting techniques which identify individuals who are more like themelves than any other individual across runs (Finn et al., 2015). Although both measures capture related measurements, connectome fingerprinting binarizes identifiability (i.e., an individual is either most like themselves or not), while connectome discriminability values are continuous. In addition to task-relevant changes, in-scanner tasks amplify individual differences in functional connectivity patterns allowing for improved identification of individuals—i.e., better connectome fingerprinting (Finn et al., 2015; Rosenberg et al., 2016; Greene et al., 2020) and, in some cases, improved behavioral predictions (Finn et al., 2017; Greene et al., 2020, but see Finn and Rosenberg 2021, Mantwill et al., 2021, and Noble et al., 2017). One possibility is that individuals who show stable, unique connectivity profiles during

task performance may be reliably engaging individualized networks supporting task performance and thus may show better task performance. Thus, we tested whether connectome discriminability predicts sustained attention and working memory performance. We observed an inconsistent relationship between discriminability and performance across datasets and tasks. Fixed effects models predicting both sustained attention and working memory were significant but less optimal than other models tested. These results suggest that connectome discriminability may not be a reliable predictor of cognitive performance, at least in the case of sustained attention and working memory. Furthermore, discriminability as defined in the current study as the ratio between one's stability and typicality did not add unique predictive power above and beyond these component measures.

Stability and typicality of individual network connections was consistent across datasets and tasks. While no network reliably related to connectome stability's prediction of performance across datasets, computational lesioning of the frontoparietal network significantly impacted the relationship between typicality and working memory performance in both datasets with a working memory measure. These results are in line with previous work that identified frontoparietal activity as a biomarker of working memory performance in preadolescents (Rosenberg et al., 2020). Inconsistent contributions from other canonical networks, in combination with limited behavioral prediction from stability within previously validated network-based models, suggest that connections involved in connectome stability and typicality's prediction of both sustained attention and working memory performance are broadly distributed. One possibility is that the canonical networks used here are too large-scale to identify reliable anatomical regions meaningful for prediction. Additionally, future studies may investigate whether the stability and typicality of functional connections relate to performance in a manner specific to the unique task being

performed. Perhaps specific task demands influence which networks' stability most contribute to performance prediction.

In the current study, we observed utility of including individual and group-related functional connectivity features in behavioral prediction models. In particular, connectome stability and optimality were able to predict sustained attention and working memory performance in datasets where network-based strength models did not generalize. Functional connectivity features have the additional benefit of being inherent to the connectome, such that models incorporating these features do not require separate training and testing sets. This suggests that unsupervised models based on functional connectivity features may capture unique variance in cognitive performance more broadly and may be valuable to include in future predictive models, particularly in cases when supervised CPMs do not generalize. The specific utility of connectome features for prediction may depend on task and dataset. While we observed significant prediction of both sustained attention and working memory using features of connectome typicality and optimality, these features are dataset-dependent such that they require meaningful patterns of connectivity to be captured in the average connectomes of the group and top performers. Alternatively, connectome stability is relatively agnostic to characteristics of the larger dataset and may therefore be a useful predictor in more idiosyncratic datasets.

**Limitations**

As is the case in many studies relating task-based functional connectivity to behavior, it is possible that our results are influenced by in-scanner task performance. For example, behavior such as motion (e.g., task-related button presses or head motion) may have systematically varied with our measures of sustained attention and working memory performance. If this were the case,

our neural measures may not be solely markers of these cognitive mechanisms but instead reflect some behavioral covariates. While the results of our behavioral stability control analysis suggest that stability in in-scanner performance does not fully explain our findings, we cannot rule out the possibility that our results are influenced by other, systematically varying behaviors.

Although the replication of effects across datasets that measured sustained attention and working memory with different tasks is a strength of the current study, that differences between tasks may have limited our findings. For example, while the 0- back task used as a measure of sustained attention in Dataset 3 resembles a target detection task, commonly used in previous literature, it likely differs in many ways from a continuous performance task such as the gradCPT used in Datasets 1 and 2. Similarly, the working memory tasks used—the VSTM in Dataset 2 and 2-back task in Dataset 3—likely involve different processes despite being used to measure the same cognitive ability. Future work may seek to investigate task-specific influences on the relationship between functional connectivity features and performance to account for these differences.

**Conclusion**

Here, we demonstrate the utility of functional connectome features as predictors of cognitive ability in adults. Of the features tested, connectome stability and optimality consistently predicted both sustained attention and working memory across three independent datasets. Stability during sustained attention and working memory tasks was related to performance on the respective task, suggesting that better performers display more similar task-specific functional connectivity patterns over time. Additionally, connectome optimality, or similarity to high performers' task connectome predicted performance across datasets, indicating the potential for

task-specific “template” patterns of connectivity. Finally, both connectome stability and optimality explained unique variance in task performance when generalizing network-based connectivity models to novel tasks. Altogether, the current study provides evidence for the utility of connectome features that summarize similarity within and across individuals as brain-based markers of cognition.

## Chapter 1 Bibliography


Amico E, Goñi J. The quest for identifiability in human functional connectomes. Sci Rep. 2018:8(1):8254. https://doi.org/10.1038/ s41598-018-25089-1.

Avery EW, Yoo K, Rosenberg MD, Greene AS, Gao S, Na DL, Scheinost D, Constable TR, Chun MM. Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. J Cogn Neurosci. 2020:32(2):241–255. https://doi.org/10.1162/jocn_a_01487.

Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, et al. Function in the human connectome: task-fMRI and individual differences in behavior. NeuroImage. 2013:80:169–189. https://doi.org/10.1016/j. neuroimage.2013.05.033.

Bassett DS, Sporns O. Network neuroscience. Nat Neurosci. 2017:20(3): 353–364. https://doi.org/10.1038/nn.4502.

Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004:33(2): 261–304. https://doi.org/10.1177/0049124104268644.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996:29(3): 162–173. https://doi.org/10.1006/cbmr.1996.0014.

Elliott ML, Knodt AR, Cooke M, Kim MJ, Melzer TR, Keenan R, Ireland D, Ramrakha S, Poulton R, Caspi A, et al. General functional connectivity: shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. NeuroImage. 2019:189:516–532. https://doi.org/10.1016/ j.neuroimage.2019.01.068.

Esterman M, Noonan SK, Rosenberg M, DeGutis J. In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. Cereb Cortex. 2013:23(11):2712–2723. https:// doi.org/10.1093/cercor/bhs261.

Finn ES, Rosenberg MD. Beyond fingerprinting: choosing predictive connectomes over reliable connectomes. NeuroImage. 2021: 239:118254. https://doi.org/10.1016/j.neuroimage.2021.118254.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci. 2015:18(11):1664–1671. https://doi. org/10.1038/nn.4135.

Finn ES, Scheinost D, Finn DM, Shen X, Papademetris X, Constable RT. Can brain state be manipulated to emphasize individual differences in functional connectivity? NeuroImage. 2017:160:140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064.

Fu Z, Liu J, Salman M, Sui J, Calhoun V. Functional connectivity uniqueness and variability? A signature of cognitive and psychiatric problems in children [Preprint]. In Review2022. https://doi. org/10.21203/rs.3.rs-1514598/v1.

Gao M, Wong CHY, Huang H, Shao R, Huang R, Chan CCH, Lee TMC. Connectome-based models can predict processing speed in older adults. NeuroImage. 2020:223:117290. https://doi.org/10.1016/j. neuroimage.2020.117290.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, et al. The minimal preprocessing pipelines for the human

connectome project. NeuroImage. 2013:80:105–124. https://doi.org/10.1016/j. neuroimage.2013.04.127.
Gonzalez-Castillo J, Hoy CW, Handwerker DA, Robinson ME, Buchanan LC, Saad ZS, Bandettini PA. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. Proc Natl Acad Sci. 2015:112(28):8762–8767. https://doi. org/10.1073/pnas.1501242112.
Gratton C, Laumann TO, Nielsen AN, Greene DJ, Gordon EM, Gilmore AW, Nelson SM, Coalson RS, Snyder AZ, Schlaggar BL, et al. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. Neuron. 2018:98(2):439–452.e5. https://doi.org/10.1016/j. neuron.2018.03.035.
Greene AS, Gao S, Noble S, Scheinost D, Constable RT. How tasks change whole-brain functional organization to reveal brain-phenotype relationships. Cell Reports. 2020:32(8):108066. https://doi.org/10.1016/j.celrep.2020.108066.
Greene AS, Gao S, Scheinost D, Constable RT. Task-induced brain state manipulation improves prediction of individual traits. Nat Commun. 2018:9(1):2807. https://doi.org/10.1038/s41467-018- 04920-3.
Hilger K, Ekman M, Fiebach CJ, Basten U. Efficient hubs in the intelligent brain: nodal efficiency of hub regions in the salience network is associated with general intelligence. Dermatol Int. 2017:60:10–25. https://doi.org/10.1016/j.intell.2016.11.001.
Kardan O, Stier AJ, Cardenas-Iniguez C, Pruin JC, Schertz KE, Deng Y, Chamberlain T, Meredith WJ, Zhang X, Bowman JE. Connectomebased predictions reveal developmental change in the functional architecture of sustained attention and working memory. BioRxiv. 2022:2021–08.
Kaufmann T, Alnæs D, Doan NT, Brandt CL, Andreassen OA, Westlye LT. Delayed stabilization and individualization in connectome development are related to psychiatric disorders. Nat Neurosci. 2017:20(4):513–515. https://doi.org/10.1038/nn.4511.
Lee AT, Glover GH, Meyer CH. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic resonance functional neuroimaging. Magn Reson Med. 1995:33(6): 745–754. https://doi.org/10.1002/mrm.1910330602.
Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997:390(6657):279–281. https:// doi.org/10.1038/36846.
Lynch LK, Lu K-H, Wen H, Zhang Y, Saykin AJ, Liu Z. Task-evoked functional connectivity does not explain functional connectivity differences between rest and task conditions. Hum Brain Mapp. 2018:39(12):4939–4948. https://doi.org/10.1002/hbm.24335.
Mantwill M, Gell M, Krohn S, Finke C. Fingerprinting and behavioural prediction rest on distinct functional systems of the human connectome. BioRxiv. 2021:2021.02.07.429922. https://doi.org/ 210.1101/021.02.07.429922.
Miranda-Dominguez O, Mills BD, Carpenter SD, Grant KA, Kroenke CD, Nigg JT, Fair DA. Connectotyping: model based fingerprinting of the functional connectome. PLoS One. 2014:9(11):e111048. https://doi.org/10.1371/journal.pone.0111048.
Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cereb Cortex. 2017:27(11):5415–5429. https://doi.org/10.1093/cercor/bhx230.
Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: a systematic review and metaanalysis. NeuroImage. 2019:203:116157. https://doi.org/10.1016/j. neuroimage.2019.116157.

Pashler H. Familiarity and visual change detection. Percept Psychophys. 1988:44(4):369–378. https://doi.org/10.3758/BF03210419.
Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, Chun MM. A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci. 2016:19(1): 165–171. https://doi.org/10.1038/nn.4179.
Rosenberg MD, Martinez SA, Rapuano KM, Conley MI, Cohen AO, Cornejo MD, Hagler DJ, Meredith WJ, Anderson KM, Wager TD, et al. Behavioral and neural signatures of working memory in childhood. J Neurosci. 2020:40(26):5090–5104. https://doi. org/10.1523/JNEUROSCI.2841-19.2020.
Rypma B, Berger JS, D'Esposito M. The inf luence of workingmemory demand and subject performance on prefrontal cortical activity. J Cogn Neurosci. 2002:14(5):721–731. https://doi. org/10.1162/08989290260138627.
Satterthwaite TD, Wolf DH, Erus G, Ruparel K, Elliott MA, Gennatas ED, Hopson R, Jackson C, Prabhakaran K, Bilker WB, et al. Functional maturation of the executive system during adolescence. J Neurosci. 2013:33(41):16249–16261. https://doi. org/10.1523/JNEUROSCI.2345-13.2013.
Schultz DH, Cole MW. Higher intelligence is associated with less task-related brain network reconfiguration. J Neurosci. 2016:36(33):8551–8561. https://doi.org/10.1523/JNEUROSCI.0358- 16.2016.
Shen X, Tokoglu F, Papademetris X, Constable RT. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. NeuroImage. 2013:82:403–415. https:// doi.org/10.1016/j.neuroimage.2013.05.081.
Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat Protoc. 2017:12(3):506–518. https://doi. org/10.1038/nprot.2016.178.
Shine JM, Bissett PG, Bell PT, Koyejo O, Balsters JH, Gorgolewski KJ, Moodie CA, Poldrack RA. The dynamics of functional brain networks: integrated network states during cognitive task performance. Neuron. 2016:92(2):544–554. https://doi.org/10.1016/j. neuron.2016.09.018.
Stanley ML, Simpson SL, Dagenbach D, Lyday RG, Burdette JH, Laurienti PJ. Changes in brain network efficiency and working memory performance in aging. PLoS One. 2015:10(4):e0123950. https://doi. org/10.1371/journal.pone.0123950.
van den Heuvel MP, Stam CJ, Kahn RS, Pol HEH. Efficiency of functional brain networks and intellectual performance. J Neurosci. 2009:29(23):7619–7624. https://doi.org/10.1523/JNEUROSCI.1443- 09.2009.
Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn human connectome project: an overview. NeuroImage. 2013:80:62–79. https://doi.org/10.1016/j. neuroimage.2013.05.041.
Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. NeuroImage. 2015:122:222–232. https:// doi.org/10.1016/j.neuroimage.2015.07.069.
Vanderwal T, Eilbott J, Kelly C, Frew SR, Woodward TS, Milham MP, Castellanos FX. Stability and similarity of the pediatric connectome as developmental measures. NeuroImage. 2021:226:117537. https://doi.org/10.1016/j.neuroimage.2020 .117537.
Xie H, Gonzalez-Castillo J, Handwerker DA, Bandettini PA, Calhoun VD, Chen G, Damaraju E, Liu X, Mitra S. Time-varying whole-brain functional network connectivity coupled to

task engagement. Network Neuroscience. 2018:3(1):49–66. https://doi.org/10.1162/netn_a_00051.
Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zôllei L, Polimeni JR, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011:106: 1125–1165.
Yoo K, Rosenberg MD, Kwon YH, Lin Q, Avery EW, Sheinost D, Constable RT, Chun MM. A brain-based general measure of attention. Nat Hum Behav. 2022a:6(6):782–795. https://doi.org/10.1038/ s41562-022-01301-1.
Yoo K, Rosenberg MD, Kwon YH, Scheinost D, Constable RT, Chun MM. A cognitive state transformation model for taskgeneral and task-specific subsystems of the brain connectome. NeuroImage. 2022b:257:119279. https://doi.org/10.1016/j.neuroimage.2022.119279.

# Chapter 2: Functional brain networks predicting sustained attention are not specific to perceptual modality

The maintenance of attention to information over time is essential for daily activities such as driving to work or conversing with friends. Recent work in cognitive neuroscience has been aimed at identifying neural signatures of sustained attention ability with the goal of constructing models that generalize across people and datasets to predict individual differences in attention function. However, while sustained attention can be deployed to information from multiple perceptual (e.g., visual and auditory) modalities, it is an open question whether predictive models generalize across modality. For example, models trained to predict performance on visual sustained attention tasks may contain modality-specific features and therefore fail to generalize or generalize poorly to capture auditory sustained attention performance. Alternatively, features may capture modality-general aspects of attention and generalize broadly. Here, we construct and test the generalizability of models trained to predict sustained attention to visual and auditory stimuli from functional connections.

Identifying brain-based markers of cognition is beneficial both for understanding associations between functional brain organization and behavior and for developing predictive models. Network neuroscience provides a framework for the identification of interpretable neural signatures of cognition (Srivastava et al., 2022). One method, connectome-based predictive modeling (CPM), identifies functional connections, or edges, between brain regions whose strength is reliably associated with phenotypes across individuals (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017). This method has identified edge networks that predict sustained attention within and across samples of individuals (Rosenberg et al., 2016, 2020; Yoo et al., 2022).

The utility of predictive models lies in their external validity, that is, generalizability across independent datasets and contexts (Poldrack et al., 2020; Rosenberg & Finn, 2022; Scheinost et al., 2019). Successful generalizability across datasets ensures a model's accuracy in the identification of relevant features as well as robustness to differences between samples. Predictive models of sustained attention constructed using CPM have demonstrated generalizability across datasets, as well as generalization to other attention tasks and attention-related symptoms (Rosenberg et al., 2016, 2018, 2020; Yoo et al., 2022). Therefore, CPM successfully captures functional networks related to attention across contexts.

Sustained attention is often measured using tasks that require continuous vigilance for the detection or discrimination of rare stimuli (Langner & Eickhoff, 2013; Mackworth, 1948). While much work has investigated sustained attention to visual stimuli, attention can be deployed to other perceptual modalities, such as audition. Previous work has shown that the ability to sustain attention to visual and auditory information is reliable within individuals, suggesting that these abilities rely on shared cognitive mechanisms (Corriveau et al., 2024; Seli et al., 2012; Terashima et al., 2021). Work using electroencephalography and fMRI data has identified neural substrates underlying detection of both visual and auditory rare targets (Katayama & Polich, 1999; Kim, 2014; Kondo et al., 2023; Linden et al., 1999; Stevens et al., 2000), further supporting a modality-general neural basis of sustained attention. However, recent work shows that neurometabolites are differentially related to auditory and visual sustained attention (Kondo et al., 2023). Further, selective attention to visual or auditory information elicits both supramodal and modality-specific neural activation patterns (Smith et al., 2010; Stevens et al., 2000), suggesting that attending to these modalities relies on distinct neural mechanisms as well. Therefore, the extent to which visual

and auditory sustained attention networks are modality-specific may depend on the extent to which they rely on modality-specific neural mechanisms.

Here, we test the degree to which predictions from functional networks of sustained attention are biased by the perceptual modality in which they are trained. We find that a network previously defined to predict visual sustained attention predicts performance across datasets and modalities. Further, we show that models trained on auditory and visual tasks are highly generalizable across perceptual modalities. Even after the removal of features identified by both visual and auditory networks, that is, modality-general features, models successfully predict cross-modality sustained attention ability. *These results* demonstrate that sustained attention relies on distributed patterns of connectivity. Additionally, they suggest that distributed patterns may be different between perceptual modality but still capture generalizable variance in sustained attention ability across datasets and modalities.

## Methods

### *Dataset 1*

The first dataset analyzed was described in detail by Kondo et al., (2022, 2023). This study was reviewed and approved by the Research Ethics and Safety Committees of Chukyo University and ATR-Promotions. Participants provided their written informed consent to participate in this study.

Participants ($N$ = 29, ages 20–35 years) were healthy Japanese adults who completed an fMRI scan consisting of two visual runs and two auditory runs of a gradual-onset continuous performance task (gradCPT; Esterman et al., 2013; Rosenberg et al., 2013; Terashima et al., 2021). Data were collected using a 3T Magnetom Prisma MRI scanner (Siemens, Munich, Germany).

Task runs were 400 s in length. A multiband echo-planar imaging sequence was used to collect 205 volumes per task run with a repetition time (TR) of 2 s. Voxels were 2 mm × 2 mm × 2 mm. The first five volumes of each run were discarded for data analysis.

The gradCPT was developed to measure sustained attention performance. In the task, stimuli gradually transition one into the next to avoid abrupt onsets. Visual runs of the gradCPT featured round, grayscale images of city (90%) and mountain (10%) scenes. Images transitioned from off to fully visible over 1.6 s such that a stimulus reached maximum visibility every 1.6 s. Images faded from peak visibility to off as presentation of the next image began. Participants were instructed to press a button for each city scene and withhold a button press for mountain scenes.

Stimuli for the auditory gradCPT were narrations from a foreign language database, excluding Japanese narrations to avoid presentation of a participant's native language. Thus, participants used acoustic clues of the stimuli, rather than semantic clues, to judge the gender of voice streams. Narrations were performed by male (90%) and female (10%) voices and gradually transitioned from one to the next using sinusoidal ramps (Terashima et al., 2021) such that a voice reached maximum presentation every 1.6 s. Participants were instructed to press a button for male voices and withhold a button press for female voices.

Because stimuli faded from one to the next, key presses were assigned to trials in an iterative manner that first assigned unambiguous presses and then assigned more ambiguous ones. Unambiguous key presses occurring in a window from 70% presented to 40% disappeared were assigned to the current trial. Key presses that occurred outside the window were assigned to adjacent trials if no responses to those trials had been made. If no response was made to either adjacent trial, the key press was attributed to the closer trial. If either trial was an infrequent trial (mountain scene, female voice), the key press was assigned to the adjacent frequent trial. Sustained

attention performance was quantified using a measure of sensitivity (*d'*), which is calculated as the normalized hit rate minus the normalized false alarm rate for each run.

*Dataset 2*

The second dataset was collected at the MRI Research Center at the University of Chicago. Study procedures were approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago. All participants provided their written informed consent prior to participation.

Participants ($N = 60$) participated in at least one session of a two-session fMRI study collected approximately 1 week apart (mean time between sessions = 10.88 days, $SD = 9.87$ days). During both sessions, participants performed a 10-min audio-visual continuous performance task (avCPT; Corriveau et al., 2024). Functional MRI data were collected on a 3T Philips Ingenia scanner. Voxels were 2.526 mm × 2.526 mm × 3 mm. Volumes were collected using a multiband sequence with a TR of 1 s. Three volumes were removed from the start of each scan.

During the avCPT, streams of trial-unique images and sounds were presented simultaneously. Images were presented continuously for 1.2 s each, whereas sounds were presented for 1 s with a 200-ms intertrial interval to allow participants to distinguish individual sounds. Each task run was 500 trials in length. Images were indoor and outdoor scenes drawn from the Scene UNderstanding image database (SUN; Xiao et al., 2010). Sound stimuli were natural and manmade sounds drawn from online sound databases and cropped to be 1 s in length. Full details of stimulus curation procedures are described in Corriveau et al., (2024).

Before the task run, participants were instructed to make a button press to frequent stimuli (90%) from either the auditory or visual modality and to withhold a button press for infrequent

stimuli (10%). They were told that the stimuli from the other modality were not relevant for the task. Over the two scan sessions, participants performed both the auditory-relevant and visual-relevant tasks, and the order of task runs and frequent stimulus category was counterbalanced across participants.

For frequent trials, correct responses were trials in which participants responded before the onset of a new stimulus (within 1,200 ms of trial start). However, to allow for the possibility of reaction times (RTs) longer than 1,200 ms, we reassigned key presses for frequent trials that met the following criteria: (a) The participant made more than one key press for a trial with a frequent-category stimulus, (b) the first key press was faster than 100 ms, and (c) no response was made to the previous frequent-category stimulus. In this case, the first key press was attributed to the previous trial. This reassignment of key presses is meant to more accurately account for correct performance with slower response times. Press reassignment was rare in both visual (mean number of trials with presses reassigned = 0.548, *SD* = 0.861) and auditory sessions (mean number of trials with presses reassigned = 3.81, *SD* = 3.46), affecting less than 0.8% of the trials in each task. Therefore, this analytical decision has a negligible effect on results. Performance during the avCPT was calculated as sensitivity (*d′*).

*Dataset 3*

The final fMRI dataset analyzed was described in Walz et al., (2013) and shared on OpenNeuro (ds000116). This dataset contained runs from 17 adults (six females, ages 20–40 years) who performed three auditory and three visual runs of an oddball task. Simultaneous fMRI and electroencephalography data were collected for the original study but only the fMRI data are analyzed here. Data were collected on a 3T Philips Achieva scanner. Voxels were 3 mm × 3 mm

× 4 mm. Each run consisted of 170 volumes collected with a 2-s TR. While the authors note that discarding of extra runs is unnecessary for the shared data, the first three volumes of each run were removed in keeping with a standard preprocessing pipeline. We do not expect this to affect the current results.

Task runs consisted of 125 stimuli presented for 200 ms with a variable intertrial interval of 2–3 s. Participants were instructed to press a button for infrequent targets (20%) and could ignore standard stimuli (80%). In visual runs, standard trials consisted of a small green circle and target trials were the presentation of a large red circle. For auditory runs, the standard stimulus was a 390-Hz tone, whereas the target stimulus was a broadband laser gun sound.

Because the response pattern for this task was inverted and responses were only required on target trials, detection of oddball targets in this task is trivial, leading to overall high performance. Therefore, sustained attention performance in this dataset was quantified using the mean run RT variability, which has previously been shown to be robustly related to sustained attention performance in both healthy adults and in populations characterized by sustained attention deficits (Chidharom, Krieg, & Bonnefond, 2021; Esterman et al., 2013; Karamacoska et al., 2018; Robertson et al., 1997; Seli et al., 2012; Tamm et al., 2012). Importantly, this measure provides more variability across participants than a measure of sensitivity on a task where performance is at ceiling, as in the current dataset. RT variability is predictive of sustained attention ability such that individuals with more variable pressing show worse performance on sustained attention tasks. Since RT variability has previously been shown to be negatively related to sustained attention performance, we report the inverse of RT variability (mean RT / standard deviation) for ease of comparison with Datasets 1 and 2 in the current study.

*fMRI Preprocessing Procedure*

Functional MRI data for the three datasets underwent the same preprocessing steps in AFNI (Cox, 1996). Preprocessing included the following steps: Removal of leading TRs as previously noted for individual datasets; alignment of functional data to MNI space; regression of covariates of no interest, including a 24-parameter head motion model (six motion parameters, six temporal derivatives, and their squares), mean signal from subject-level white matter and ventricle masks, and mean whole-brain signal; and censoring of volumes for which the derivative of motion parameters exceeded 0.25 mm or for which more than 10% of the brain were outliers.

*Exclusion Criteria*

To ensure high-quality data, individual runs were excluded if they did not meet the following criteria regarding head motion inside the scanner and behavioral performance. Runs were excluded if mean framewise head displacement after motion censoring exceeded 0.15 mm, if the maximum head displacement exceeded 4 mm, or if greater than 50% of frames were censored during preprocessing. Runs in Datasets 1 and 2 were also excluded if hit rates were more than 2.5 standard deviations below the mean hit rate value. The tasks used in these datasets asked participants to respond to frequent trials (90%), such that good performance would require presses to the vast majority of trials. Therefore, low hit rates for these tasks indicate participant noncompliance. Finally, we excluded runs if behavioral performance, quantified as sensitivity ($d'$) in Datasets 1 and 2 and inverse RT variability in Dataset 3, was greater than 2.5 standard deviations below the mean across all runs within a dataset.

In Dataset 1, two visual runs were removed based on head motion criteria and six visual runs were excluded for extremely low hit rates. No auditory runs were removed based on any of

the listed criteria. In Dataset 2, 56 participants completed the visual avCPT and 55 participants completed the auditory avCPT. Nine visual runs and 10 auditory runs were excluded based on head motion criteria. An additional two visual runs and one auditory run were removed due to low hit rates. In the final sample for Dataset 2, 36 participants completed both a visual and an auditory run. For Dataset 3, 47 visual and 44 auditory runs were successfully preprocessed. Preprocessing failed for the remaining four visual and seven auditory runs due to the number of time points censored. No additional runs were removed based on head motion criteria. No runs in any dataset were excluded on the basis of low sensitivity or RT variability measures. The final sample sizes for each dataset and run type were as follows: Dataset 1 included 50 visual and 58 auditory runs, Dataset 2 included 45 visual and 44 auditory runs, and Dataset 3 included 47 visual and 44 auditory runs.

*External Validation of Sustained Attention CPM*

Functional MRI data were parcellated into 268 functionally defined regions of interest (ROIs; Shen et al., 2013). Whole-brain functional connectivity matrices, or functional connectomes, were calculated by correlating the blood oxygen level-dependent (BOLD) time courses for a given task run between all pairs of ROIs. Edges in this 268-by-268 matrix provide an index of coactivation similarity between all pairs of regions in the brain for each run.

Our first question of interest was whether a predefined network trained to predict sustained attention performance in a visual task generalized to the present datasets, which include both visual and auditory sustained attention tasks. The network tested was defined using CPM (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017), which identifies a set of edges whose coactivation strength is related to a performance metric across a set of participants. In CPM, the strength of

every edge in a functional connectivity matrix is correlated with a behavior of interest, in this case, sustained attention performance. The predefined network, referred to in the current manuscript as the saCPM (sustained attention CPM), consists of a set of edges whose strength was either positively (757 edges) or negatively (630 edges) correlated with visual gradCPT performance across an independent set of participants ($N = 25$). Significant edges were defined as those whose network strength was significantly correlated (Pearson's $r$; $p < 0.01$) with visual gradCPT sensitivity ($d'$) across participants. Positively correlated edges are connections whose strength increased with higher sustained attention performance across participants, whereas negatively correlated edges are connections whose strength increased with worse performance. This network is described in previous work by Rosenberg et al., (2016, 2020) and is shared publicly (https://github.com/monicadrosenberg/Rosenberg_PNAS2020).

Here, we tested whether strength in this predefined visual network also predicted sustained attention performance in datasets that include novel participants, multiple perceptual modalities, and new behavioral measures of interest. Network strength is defined as the difference between mean connectivity in the high-attention network and mean connectivity in the low-attention network for each run in the current datasets. Because strength in the high-attention and low-attention networks will be negatively correlated by nature of how the networks were identified, taking the difference provides a single summary measure that is interpretable. CPM-predicted behavior is a linear transformation of network strength (predicted behavior = $m \times$ network strength + $b$, where $m$ and $b$ are learned during model training). Therefore, for external model validation as we perform in the current set of analyses, the correlation between network strength and observed behavior is mathematically equivalent to correlation between predicted and observed behavioral scores. Network strength values were normalized across participants within dataset for comparison

with other analyses. We then tested whether network strength was related to behavioral performance by calculating the partial Spearman's *rho* value between network strength and the behavioral measure of interest for visual and auditory runs separately, controlling for mean head motion (mean framewise displacement) in the scanner. Spearman's rho values were used to mitigate any potential effects of outliers on predictions. However, results are consistent when using Pearson's correlation.

As a note, we do not apply multiple comparison correction for the present study because all tests of model generalization tested a nonomnibus hypothesis, that is, that network strength in the trained model will predict sustained attention performance in an independent sample (García-Pérez, 2023). Each external validation of model prediction tests a single outcome (significance of correlation between network strength and performance) and therefore multiple comparisons corrections would create unnecessarily large barriers to generalization.

*Modality-Specific Model Construction*

Next, we tested whether a network that is trained on fMRI data collected during a sustained attention task performed in a given perceptual modality better predicts performance on a task performed in the same versus a different modality. To test this, we defined new models on the functional connectivity matrices and behavior in Dataset 1 using a CPM approach (Finn et al., 2015; Rosenberg et al., 2016; Shen et al., 2017). CPM identifies a set of edges that is correlated, either negatively or positively (Pearson's $r$, $p < 0.01$) with behavioral performance across the training set. For the current analyses, the training set was all visual or auditory runs in Dataset 1. For each edge in a functional connectome a Pearson's correlation is calculated between edge strength and sustained attention performance across the dataset. This is repeated for all edges in the functional connectivity matrix, and significant edges are those whose correlation with sustained

attention is stronger than a given threshold, in this case, $p < 0.01$. Positive network edges are those where connectivity strength is positively related to behavior across an entire training sample, while negative network edges are those whose connectivity is negatively related to behavior across the sample. Significant edges are isolated to represent a network of edges for which edge strength is related to sustained attention in a given dataset. This results in binary edge "masks" consisting of 0 s and 1 s for both positive and negative networks. Edge masks are used to calculate network strength in independent datasets by calculating the dot product between the binary mask and each individual's functional connectivity matrix and taking the difference between average connectivity strengths in the positive and negative edge networks. Networks were defined on visual and auditory runs of Dataset 1 separately. We then tested the generalizability of these networks by calculating the partial Spearman's correlation between modality-specific network strength and performance in the left-out Datasets 2 and 3 visual and auditory runs, controlling for in-scanner head motion. For these external validation analyses, the correlation between network strength and observed behavioral scores is again equivalent to the correlation between predicted and observed behavioral scores.

To investigate the composition of visual, auditory, and overlapping sustained attention networks, we quantified the relative contribution of canonical brain networks (Finn et al., 2015) to these networks. This functionally defined canonical network parcellation includes visual networks labeled based on their similarity to resting-state visual networks. There is no comparable auditory network included in this parcellation. However, connections from the auditory cortex may be best encompassed by medial frontal and motor networks. We quantified relative contribution to high and low sustained attention networks by calculating the difference between the number of edges identified by high and low networks within and between canonical networks. This relative

contribution was normalized by the total number of edges contained in a network. Significance of network contributions was calculated nonparametrically by shuffling edges in the high and low attention networks separately and recalculating the difference in network contribution 1,000 times.

We quantified the significance of overlap between our new visual and auditory sustained attention networks using a hypergeometric cumulative distribution function, which calculates the probability of observing the number of overlapping edges given a random sampling with no replacement of two networks of the sizes observed (Rosenberg et al., 2016). Significance values were calculated in MATLAB as $p = 1\text{-}hygcdf(x, M, K, N)$ where $x$ is the number of shared edges between networks of interest, $M$ is the total number of functional edges in the matrix (35,778), and $K$ and $N$ are the number of functional edges the networks of interest.

We tested whether model generalization was biased toward the perceptual modality of training by calculating a measure of modality specificity for each external validation dataset. Modality specificity of visual and auditory networks was calculated as the prediction (partial Spearman's rho) of within-modality generalization (e.g., visual performance predicted by the visual CPM) minus cross-modality generalization (e.g., visual performance predicted by the auditory CPM) for each dataset and modality. We determined significance with a permutation test whereby predicted performance values were shuffled and correlated with observed performance, controlling for head motion. Auditory and visual predicted performance values were shuffled independently, and the difference between these partial Spearman's rho values was calculated. This process was repeated across 5,000 iterations to obtain a null distribution of permuted difference scores.

Finally, we tested the contribution of network components to the generalizability of auditory and visual networks. To do so, we calculated whether network strength in reliably

predictive edges, that is, edges that appeared in both visual and auditory predictive networks, was related to sustained attention performance in independent datasets. We hypothesized that these edges would reflect connectivity involved in supramodal sustained attention and therefore would generalize to predict performance in both modalities. We also tested whether edges that appeared only in the visual network or the auditory network would show specificity to their training modality. To do this, we calculated network strength in edges that appeared either in the visual network or the auditory network, but not in both. We calculated the modality specificity of visual-only and auditory-only network edges by comparing predictions within and across modality, as described in the previous paragraph.

All preprocessed data and analysis code required to recreate the described analyses are publicly available at https://osf.io/bt2xy/.

## Results

### *A Predefined Visual Network Generalizes Across Datasets and Modalities*

We first tested whether the saCPM, a network trained to predict performance on a visual sustained attention task generalized to the current datasets. Sustained attention performance, measured as sensitivity ($d'$) in Dataset 1, ranged from 1.14 to 5.11 in visual runs ($M = 3.11$, $SD = 0.957$) and 0.306 to 3.69 in auditory runs ($M = 1.51$, $SD = 0.679$). In Dataset 2, visual $d'$ values ranged from 1.08 to 4.40 ($M = 3.09$, $SD = 0.640$) and auditory $d'$ values ranged from −0.112 to 2.22 ($M = 0.968$, $SD = 0.599$). Inverse RT variability in Dataset 3 ranged from 3.67 to 10.88 ($M = 7.39$, $SD = 1.76$) in visual runs and from 2.76 to 11.47 ($M = 5.53$, $SD = 1.98$) in auditory runs. Mean visual and auditory sustained attention measures were positively related across subjects in all datasets (Spearman's $rho_1 = 0.497$, $p_1 = 9.84 \times 10^{-3}$; Spearman's $rho_2 = 0.537$, $p_2 < 0.001$;

Spearman's $rho_3 = 0.589$, $p_3 = 0.021$). However, performance was not perfectly correlated across modalities, such that not all variance in auditory task performance was explained by performance on the visual task. Therefore, successful generalization of the saCPM would require that it rely on features which capture shared supramodal variance.

For visual task runs, network strength in the saCPM was positively related to performance in all datasets (partial $rho_1 = 0.230$, $p_1 = 0.112$; partial $rho_2 = 0.317$, $p_2 = 0.036$; partial $rho_3 = 0.343$, $p_1 = 0.020$), and this relationship was significant in Datasets 2 and 3 (**Figure 7**). While the prediction of visual sustained attention was not significant in Dataset 1, the relationship between network strength and observed performance was in the predicted direction and aligns with predictions in other datasets. As a validation that the saCPM captures visual sustained attention performance in Dataset 1, we also tested whether network strength in the saCPM predicted inverse RT variability in this dataset. RT variability was significantly correlated with sustained attention performance in visual runs of Dataset 1 ($r = 0.649$, $p < 0.001$) but may capture more meaningful variance in performance in this dataset. saCPM network strength positively predicted inverse RT variability during the visual task (partial $rho = 0.313$, $p = 0.028$). Therefore, we concluded that this previously validated network generalizes to predict visual sustained attention performance in Dataset 1. When applied to auditory task runs, saCPM network strength significantly predicted auditory sustained attention performance in all three datasets (partial $rho_1 = 0.344$, $p_1 = 8.74 \times 10^{-3}$; partial $rho_2 = 0.360$, $p_2 = 0.018$; partial $rho_3 = 0.552$, $p_1 < 0.001$). Successful generalization of the predefined saCPM demonstrates that this network captures features of sustained attention that are general across datasets as well as perceptual modalities.

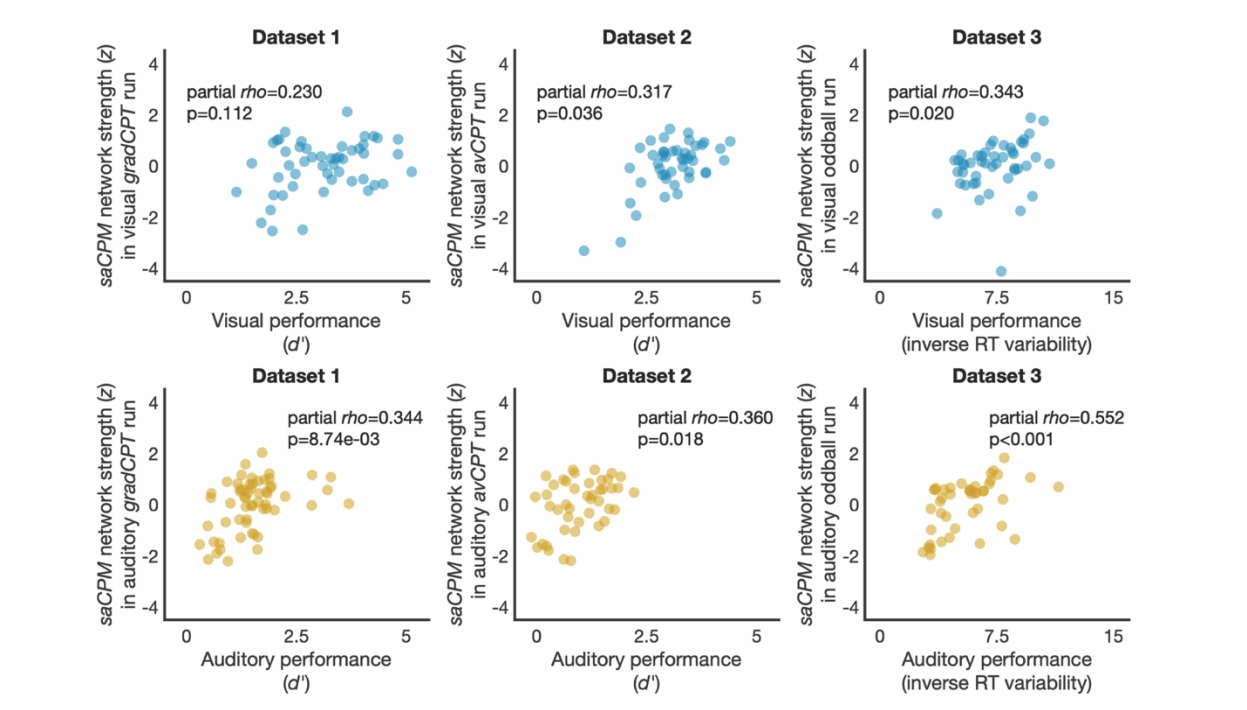



*Figure 7. The sustained attention CPM predicts visual and auditory performance.* Network strength in the saCPM network significantly predicted visual sustained attention performance in Datasets 2 and 3 and auditory sustained attention performance in all datasets.

*Sustained Attention Networks Are Not Modality Specific*

We next asked whether a model trained on an auditory sustained attention task would generalize to predict performance on other auditory attention tasks better than a model trained on a visual attention task. One option for doing so would be training a new CPM to predict auditory task performance, applying the model to new data, comparing its predictive power with that of the saCPM. However, in this scenario, any differences in predictive power could be due to differences between training datasets (number of participations, amount and quality of data, etc.) rather than attention modality per se. Thus, to more directly compare the generalizability of

auditory and visual attention models, we constructed two new models: an auditory model trained to predict auditory gradCPT performance in Dataset 1 and a visual model trained to predict visual gradCPT performance in Dataset 1. Dataset 1 was selected as the training dataset because sustained attention performance in this dataset was measured using the gradCPT. Thus, networks from this dataset are most comparable with saCPM networks, which were trained using the same task. We tested the generalizability of these models within and across perceptual modalities by relating network strength in the visual and auditory networks to visual and auditory sustained attention performance in Datasets 2 and 3, controlling for head motion during task runs. For all results reported below, models were applied to functional connectivity data from an auditory or visual task run and resulting predictions were related to behavioral performance from that same task run.

The visual network generalized to predict visual sustained attention in Dataset 2 (partial $rho = 0.329$, $p = 0.029$) and Dataset 3 (partial $rho = 0.305$, $p = 0.039$; **Figure 8A**). The auditory network similarly generalized to predict auditory sustained attention performance in both datasets (partial $rho_2 = 0.376$, $p_2 = 0.013$; partial $rho_3 = 0.537$, $p_3 < 0.001$). Within-modality generalization confirms that CPM successfully identified networks whose strength predicts out-of-sample sustained attention performance.

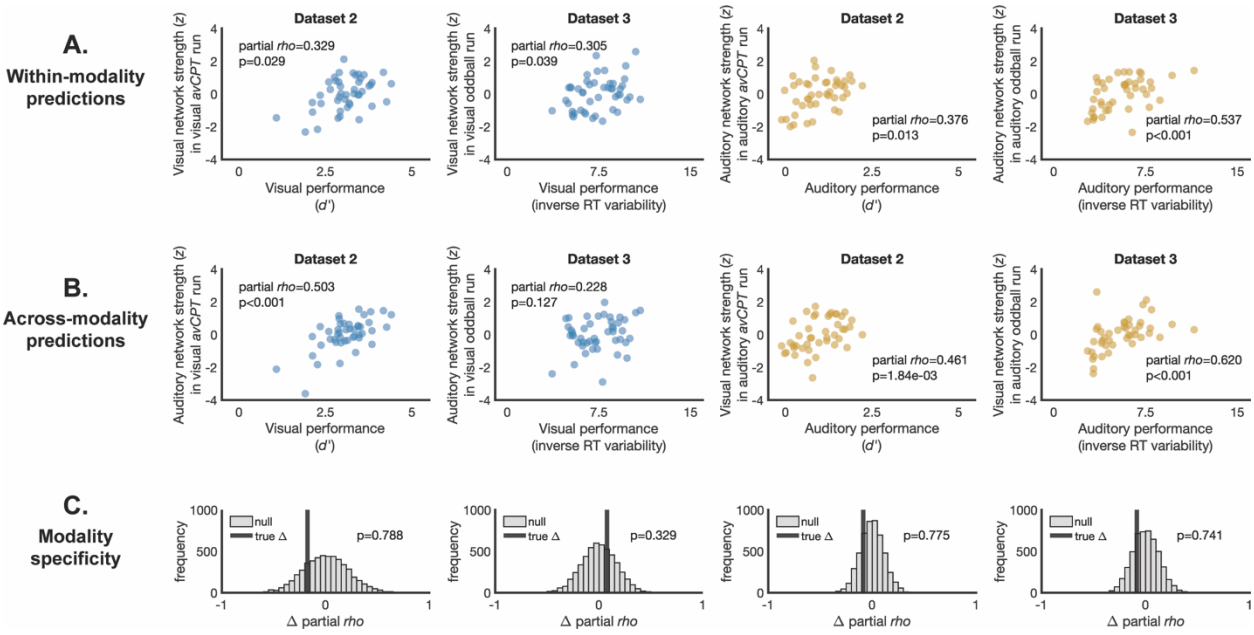



*Figure 8. Sustained attention networks are not specific to task modality.* (A) Visual and auditory networks generalized to predict visual and auditory sustained attention performance, respectively, in independent datasets. Network strength is quantified as the difference between the average high and the average low network strength values. (B) The visual network predicted auditory sustained attention in independent datasets and the auditory network predicted visual performance in one dataset. (C) Neither network showed modality specificity, that is, generalized better to within-modality prediction than across-modality prediction. The vertical black bar represents the true difference between predictions of task performance from a within-modality model versus an across-modality model. The gray distribution reflects null differences from predictions of shuffled sustained attention performance. Positive partial rho difference values reflect better prediction within versus across perceptual modality. Negative partial rho difference values reflect better performance prediction for a task performed in a different perceptual modality than training.

We next tested whether these networks predicted sustained attention performance when tasks were performed in a different modality. The visual network significantly predicted auditory sustained attention performance in both Dataset 2 (partial *rho* = 0.461, $p = 1.84 \times 10^{-3}$) and Dataset 3 (partial *rho* = 0.620, $p < 0.001$; **Figure 8B**). The auditory network predicted visual sustained attention performance in Dataset 2 (partial *rho* = 0.503, $p < 0.001$) and was positively but not significantly related to visual sustained attention performance in Dataset 3 (partial *rho* = 0.228, $p = 0.127$). Successful generalization across dataset and perceptual modality suggests that sustained attention relies on a modality-general mechanism captured, at least to some extent, by the edges identified by CPM.

To quantify the extent to which networks were modality specific, we calculated the difference between within-modality prediction and across-modality prediction by subtracting the respective partial Spearman's rho values. We created a null distribution by shuffling model-predicted visual and auditory behavioral performance values independently. The difference between partial Spearman's rho values was permuted and this was repeated 5,000 times. A one-tailed test was used to determine whether the observed difference between within- and between-modality predictions was greater than permuted values. Visual performance was not better predicted by a visual network than an auditory network in Dataset 2 ($p = 0.788$) or in Dataset 3 ($p = 0.329$; **Figure 8C**). Similarly, auditory performance predictions from the auditory network were not higher than predictions from the visual network in either Dataset 2 ($p = 0.775$) or Dataset 3 ($p = 0.741$). Across both models, we found no modality specificity such that the networks identified to predict visual or auditory sustained attention performance did not better predict task performance in the same modality.

We note that the number of runs available in Dataset 1 to train these models differs between auditory ($N = 58$) and visual ($N = 50$) runs. To ensure that training the auditory model on a larger number of runs did not bias the model's generalizability, we subsampled the number of runs used to train the auditory network to be equal to the number of runs used to train the visual network, that is, 50 runs. We refit 1,000 auditory models using a random subsampling of 50 auditory runs from Dataset 1 and tested how well these models generalized across datasets and modalities. In all cases, prediction from the model trained on the full $N = 58$ sample fell within one standard deviation of the mean prediction from models trained on a smaller sample (Prediction of visual performance: mean partial $rho_2 = 0.478$, $SD_2 = 0.054$; mean partial $rho_3 = 0.220$, $SD_3 = 0.053$; Prediction of auditory performance: mean partial $rho_2 = 0.365$, $SD_2 = 0.026$; mean partial $rho_3 = 0.534$, $SD_3 = 0.040$). Therefore, it is not the case that prediction from the auditory model in the current analyses is biased due to a larger amount of training data.

*Cross-Modality Generalization Is Not Explained by Within-Modality Performance*

Sustained attention performance is reliable across modalities, such that individuals with high visual sustained attention performance tend to have high auditory sustained attention performance. Therefore, cross-modality generalization of network predictions could result simply because cross-modality performance is related to within-modality performance. Another alternative, however, is that network models capture variance above and beyond sustained attention performance consistency. To test this, we included within-modality sustained attention performance as an additional variable in the partial correlation between cross-modality performance and network strength. If models fail to generalize when supramodal sustained attention performance is captured by the additional variable of within-modality performance, this

suggests that the generalizability of these models across modalities relies heavily on features related to this shared supramodal ability. If models still generalize after controlling for supramodal sustained attention performance, this would suggest that networks capture unique variance beyond what can be explained by similarity in sustained attention performance across runs.

The partial correlation between saCPM network strength and auditory sustained attention performance remained significant in both Dataset 1 (partial $rho = 0.301$, $p = 0.024$) and Dataset 3 (partial $rho = 0.540$, $p < 0.001$), even when controlling for participants' visual sustained attention performance. Prediction in Dataset 2 was positive but not significant after controlling for visual sustained attention performance (partial $rho = 0.215$, $p = 0.183$). Therefore, in two of three datasets, generalization of the saCPM to auditory tasks cannot be explained by a correlation between performance across modalities.

We further tested whether generalization of visual and auditory networks trained on Dataset 1 remained after controlling for within-modality performance. Predictions of auditory sustained attention performance from the visual network remained significant after removing variance explained by visual sustained attention performance in Dataset 2 (partial $rho = 0.339$, $p = 0.033$) and Dataset 3 (partial $rho = 0.605$, $p < 0.001$). When controlling for auditory sustained attention performance, predictions of visual sustained attention from auditory networks were significant in Dataset 2 (partial $rho = 0.455$, $p = 2.81 \times 10^{-3}$) and remained nonsignificant in Dataset 3 (partial $rho = 0.096$, $p = 0.532$). Therefore, generalization of sustained attention networks across task modalities is not simply due to performance similarity across modalities, but rather networks capture sustained attention ability beyond what can be explained by shared supramodal variance.

As a final test of the extent to which cross-modality generalization relies on shared variance in task performance between modalities, we retrained sustained attention networks to predict auditory and visual performance in Dataset 1, controlling for performance in the other modality. In other words, during the feature selection step of visual sustained attention model training, positive and negative network edges were those that were significantly correlated with visual sustained attention performance across individuals in Dataset 1 partialing out auditory sustained attention performance. Similarly, visual sustained attention performance was included as a partial covariate when identifying auditory sustained attention model features. Therefore, these models should no longer capture variance that can be explained by consistency in performance across individuals. A failure of these models to generalize across datasets and modalities would suggest that previous model generalization relied heavily on the shared variance in sustained attention performance across task modality. However, if these models indeed predict sustained attention performance in a modality different than training, it suggests that model features capture relevant variance beyond what can be explained by consistency in performance across modalities.

Cross-modality predictions of visual sustained attention performance from the retrained auditory network were significant in Dataset 2 (partial $rho = 0.546, p < 0.001$) and remained nonsignificant in Dataset 3 (partial $rho = 0.092, p = 0.544$). The retrained visual network significantly predicted auditory sustained attention performance in both Dataset 2 (partial $rho = 0.379, p = 0.012$) and Dataset 3 (partial $rho = 0.494, p < 0.001$). The strength of prediction, that is, partial *rho* values, were reduced in three of these cross-modal generalizations, suggesting that shared variance at least partially contributed to the generalizability of sustained attention networks. However, models' ability to significantly generalize across task modality after controlling for the

shared variance in task performance suggests that these models do not rely only on this supramodal variance.

*Unique Features Underlie Auditory and Visual Networks*

Is successful cross-modal prediction a result of shared network edges between visual and auditory networks? If CPM identified a largely overlapping set of edges related to performance on both visual and auditory tasks, it should follow that predictions would not be modality specific. However, if auditory and visual networks are independent, generalization across modalities might suggest that sustained attention performance can be captured by a diverse set of features.

The visual network consisted of 581 positive (high attention) edges and 659 negative (low attention) edges. In the auditory network, 626 edges were positively related to auditory sustained attention performance and 970 edges were negatively related to auditory sustained attention performance. Edgewise contributions to individual networks are grouped into lobe and canonical network groupings (Finn et al., 2015) and visualized in **Figure 9**.

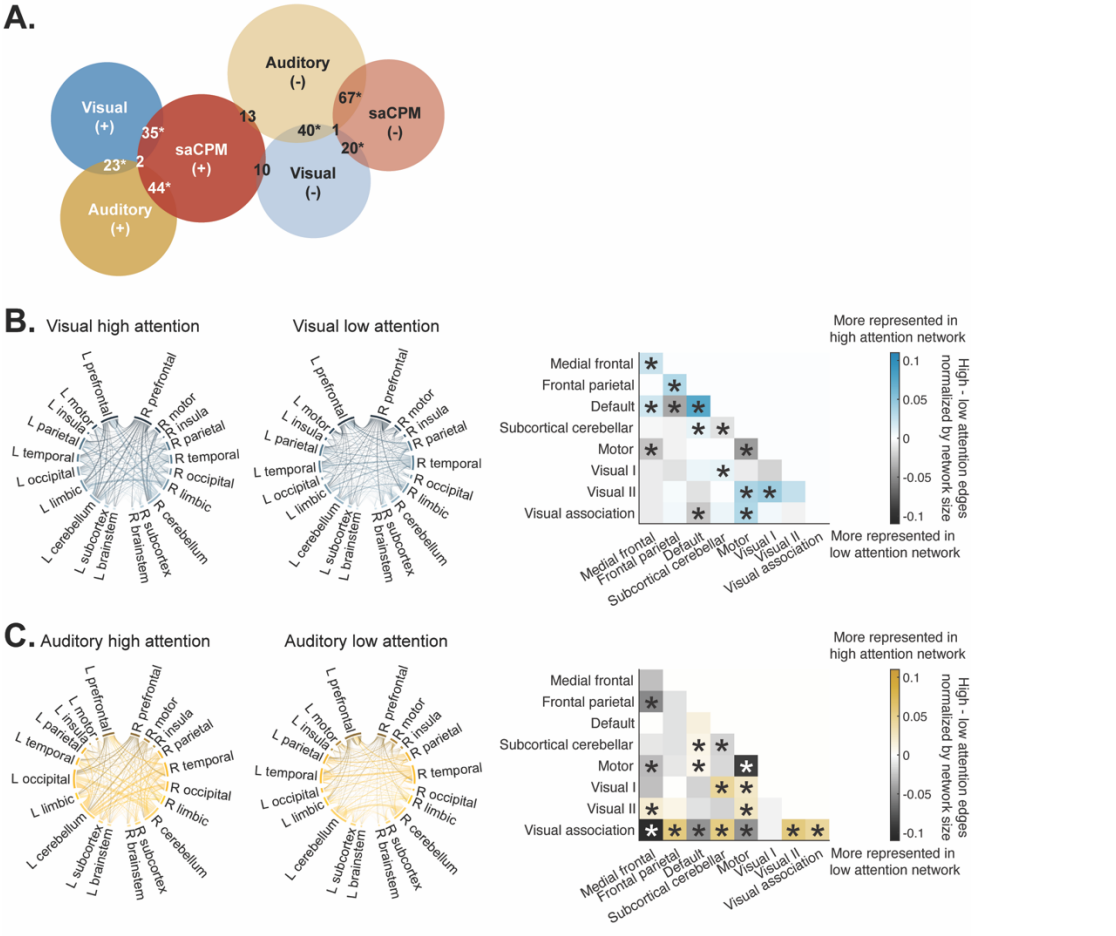



*Figure 9. Network overlap and contributions from canonical resting state networks.* (A) Networks constructed using connectome-based predictive modeling identified shared edges relevant for brain-behavior predictions in both high (+)- and low (−)-attention networks. Not all overlap between could be visualized in the Venn diagram but is described fully in the text. Stars reflect a significant number of overlapping edges between networks, $p < 0.01$. Contributions to network structure grouped by lobe and canonical network differed between (B) visual and (C) auditory networks. Matrices visualize the relative contribution of canonical network edges to high- and low-attention networks. Colors represent the difference between the number of edges in the high- and low-predictive networks, divided by network size. Stars in the matrix reflect significant contribution to high- or low-attention networks; $p < 0.05$, uncorrected. Significance was

determined by shuffling high- and low-attention networks and recalculating the contribution of edges to each network 1,000 times.

Within-network connections in the medial frontal, frontal parietal, and default mode networks contributed to the visual high-attention network (**Figure 9B**). Connections between the motor network and the visual II and visual association networks also contributed to the high-attention network, as well as connections between the default mode and medial frontal and subcortical-cerebellar networks. Within-network edges in the motor and subcortical-cerebellar networks contributed to the visual low-attention network. Connections between the default mode and frontal parietal and visual association networks were also stronger in the visual low-attention network, as well as connections between motor and medial frontal networks.

Connections within the visual association network, as well as connections between the visual association and frontal parietal, subcortical-cerebellar, and visual II networks were represented in the auditory high-attention network (**Figure 9C**). Edges shared between motor networks and visual I, visual II, and default mode networks were also represented in the auditory high-attention network. Connections between the subcortical-cerebellar network and visual I and default mode networks contributed significantly to the auditory high-attention network, as well as connections between the medial frontal and visual II networks. Conversely, connections between the visual association network and medial frontal, default mode, and motor networks were strongly represented in the auditory low-attention network. Connections between the medial frontal network and frontal parietal and motor networks were also found more strongly in the auditory low-attention network. Finally, connections within the subcortical-cerebellar networks and motor networks contributed to the auditory low-attention network.

Overlap between visual and auditory networks was significant for both high-attention (25 edges, $p < 0.001$) and low-attention networks (41 edges; $p < 0.001$; **Figure 9A**). Networks also overlapped with the predefined saCPM. Auditory networks significantly overlapped with the saCPM, sharing 46 high-attention ($p < 0.001$) and 68 low-attention edges ($p < 0.001$). The visual network also significantly overlapped with the saCPM, sharing 37 high- ($p < 0.001$) and 21 low-attention edges ($p = 3.52 \times 10^{-3}$). Network overlap in unexpected directions was nonsignificant in all cases (all $p$s > 0.826; **Figure 9A**).

*Predictions From Nonoverlapping Features Generalize Across Modality*

Does removing overlapping edges from visual and auditory networks induce modality specificity? It is possible that the generalizability of networks across modality is driven by the subset of shared edges between networks. To ask this, we tested whether edges that were unique to the auditory or visual network—for example, edges that positively predicted auditory performance but did not predict visual performance—generalized in a modality-specific manner.

Predictions of visual sustained attention performance from visual-unique model edges were significant in Dataset 2 (partial *rho* = 0.326, $p = 0.031$) and positive but nonsignificant in Dataset 3 (partial *rho* = 0.277, $p = 0.063$). Edges specific to the visual network remained generalizable across perceptual modality such that they predicted auditory sustained attention performance in both Dataset 2 (partial *rho* = 0.466, $p = 1.64 \times 10^{-3}$) and Dataset 3 (partial *rho* = 0.582, $p < 0.001$). We observed no evidence for better prediction for visual sustained attention from a model trained on a visual sustained attention task, even after removing modality-general features ($p_2 = 0.805$; $p_3 = 0.365$).

Auditory-unique edges significantly predicted auditory sustained attention performance in both Dataset 2 (partial *rho* = 0.371, $p = 0.014$) and Dataset 3 (partial *rho* = 0.537, $p < 0.001$). For visual performance, the auditory-unique network predictions were significant in Dataset 2 (partial *rho* = 0.511, $p < 0.001$) and positive but not significant in Dataset 3 (partial *rho* = 0.212, $p = 0.158$). Again, there was no evidence for modality specificity in predictions from auditory-only network edges ($p_2 = 0.783$; $p_3 = 0.606$). Therefore, predictions from nonoverlapping edges did not result in modality-specific generalization. Instead, even network edges unique to a network trained on one modality captured sustained attention ability in another modality.

Overlapping Features Are Sufficient for Prediction

Within-network edges in the default mode network, as well as edges shared between the default mode and medial frontal networks contributed to the overlapping high-attention network (**Figure 10A**). Additionally, connections shared by the frontal parietal and visual II networks, as well as connections shared between the visual association and subcortical-cerebellar networks were strongly represented in the overlapping high-attention network. This suggests that stronger connections between these networks are associated with higher modality-general sustained attention performance. Conversely, within-network edges in the visual association, subcortical-cerebellar, and motor networks contributed strongly to the overlapping low attention network. These results suggest that strong within-network connectivity in these networks is associated with worse sustained attention.

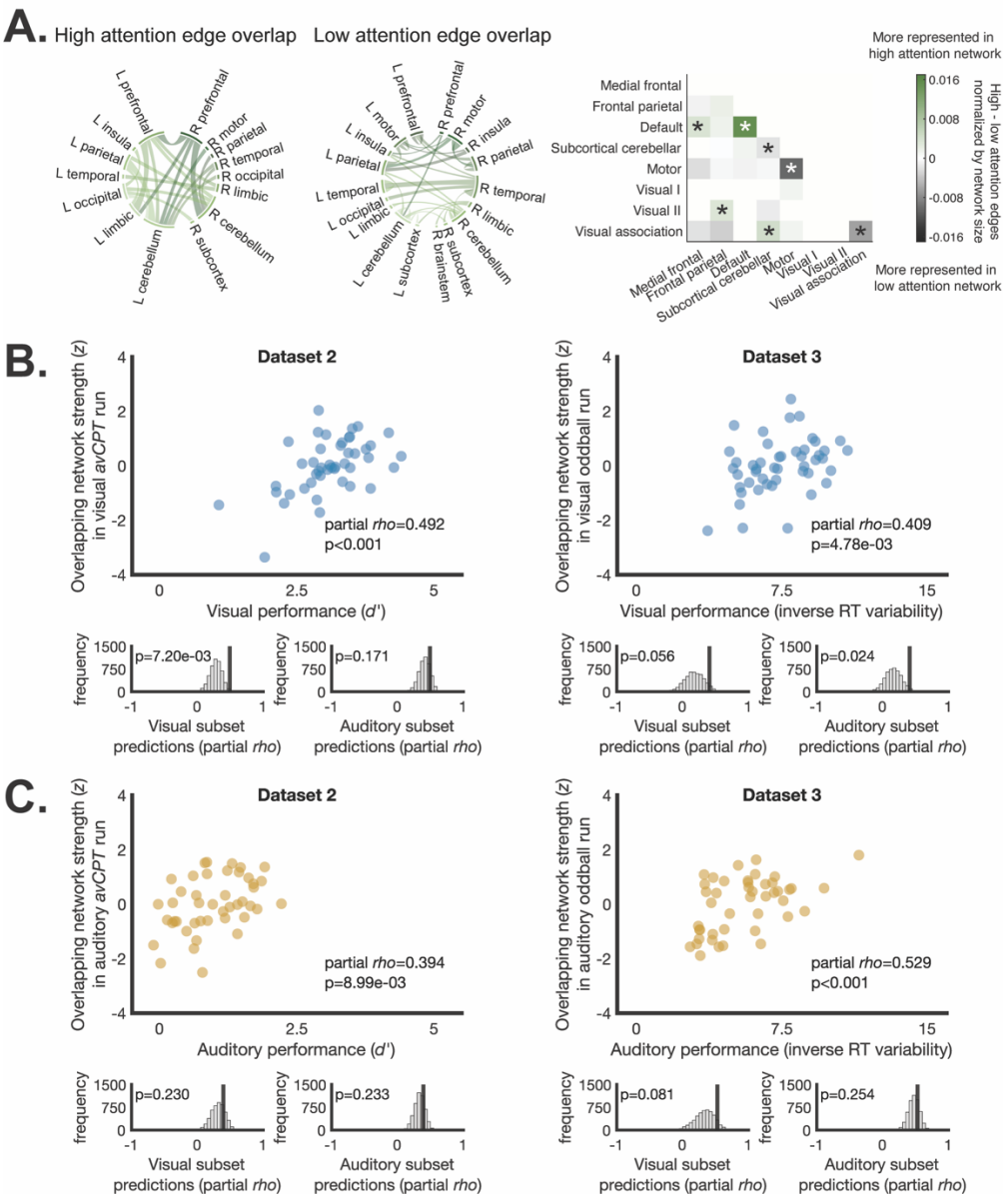



*Figure 10. Modality-general sustained attention networks generalize across datasets*. (A) Edges shared by both auditory and visual models trained on Dataset 1 are visualized by lobe. The matrix depicts relative contribution to high and low overlapping predictive networks, grouped into eight canonical networks. Significance stars on the matrix reflect greater representation of network edges than chance; $p < 0.05$, uncorrected. Significance was determined by shuffling network edges 1,000 times and recalculating relative contribution to high- and low-attention networks. Network

strength in edges shared by the auditory and visual networks predicted (B) visual and (C) auditory sustained attention performance in independent datasets. Histograms beneath each plot depict the extent to which prediction from network strength in the overlapping edges outperforms predictions from equally sized subsets of edges drawn from either the visual or auditory networks alone (5,000 permutations).

We wondered whether the edges shared by both visual and auditory networks defined in Dataset 1 were sufficient to predict visual and auditory sustained attention performance in independent datasets. To ask this, we tested whether strength in the edges shared by high visual and auditory attention networks (25 edges) and low visual and auditory attention networks (41 edges) was related to observed behavioral performance. Results are visualized in **Figure 10B–C**.

We observed robust prediction from these overlapping edges such that network strength in this subset of edges significantly predicted visual performance in Dataset 2 (partial $rho = 0.492$, $p < 0.001$) and Dataset 3 (partial $rho = 0.409$, $p = 4.78 \times 10^{-3}$), as well as auditory sustained attention performance in Dataset 2 (partial $rho = 0.394$, $p = 8.99 \times 10^{-3}$) and Dataset 3 (partial $rho = 0.529$, $p < 0.001$). Therefore, while the number of shared network features was small between visual and auditory networks, these shared features were sufficient for generalizable prediction of sustained attention performance.

Is prediction from a small set of edges specific to the shared edges between visual and auditory networks? To ask this question, we compared the predictive performance from overlapping edges with predictions from equal-sized subsets of visual or auditory network edges. Five thousand random subsets were drawn from either the visual or auditory network and network

strength from these subsets was related to observed performance. Distributions of partial *rho* values from edge subsets were created for visual and auditory networks separately.

In visual runs, edges that overlapped between the visual and auditory networks outperformed visual sustained attention performance prediction from other edge subsets of the same size drawn from the visual network in Dataset 2 ($p = 7.20 \times 10^{-3}$) but did not significantly outperform visual network edge subsets in Dataset 3 ($p = 0.056$). Predictions for visual sustained attention performance from overlapping edges did not outperform edge subsets drawn from the auditory network in Dataset 2 ($p = 0.171$) but overlap predictions did outperform auditory subset predictions and Dataset 3 ($p = 0.024$). This suggests that overlapping visual and auditory predictive edges may carry unique predictive ability related to visual sustained attention performance. When predicting auditory sustained attention performance, predictions from overlapping edges were not better than predictions from edge subsets drawn from either the visual ($p_2 = 0.230$; $p_3 = 0.081$) or the auditory network ($p_2 = 0.233$; $p_3 = 0.254$). Therefore, reliable edges provided specific predictive boost in visual runs only.

**Discussion**

Prior predictions from connectome-based models of sustained attention may have been more limited than previously thought if they were driven by visual task performance specifically. Here, we tested the extent to which functional networks of sustained attention are modality specific, with two likely outcomes. First, functional networks or a subset of functional networks may have predicted sustained attention in a modality-specific manner, generalizing better to tasks performed in the same perceptual modality as training. Alternatively, functional networks may not have shown modality specificity and predicted sustained attention performance for tasks

performed in different modalities similarly. Results show evidence for the latter, demonstrating wide-spread cross-modality generalization even when predictive model features were largely unique. This suggests that sustained attention performance can be captured by distributed, supramodal connections in the brain. Further, we demonstrated that both shared and unique edges in visual and auditory networks predict sustained attention performance across modality, showing that both reliable (overlapping) and unreliable (unique) model features capture relevant brain-behavior relationships.

Work investigating brain-behavior relationships emphasizes that testing model generalizability, in particular, generalizability to external datasets, is the gold standard for the construction of accurate predictive models (Poldrack et al., 2020; Rosenberg & Finn, 2022; Scheinost et al., 2019). Connectome-based predictive models have previously demonstrated robust generalizability to predict relevant cognitive phenotypes across independent samples (Avery et al., 2020; Fountain-Zaragoza et al., 2019; Gao et al., 2020; Kardan et al., 2022; Rosenberg et al., 2016, 2018, 2020). Therefore, CPM meets this high benchmark for model validity and holds promise for identifying robust and interpretable predictors of cognitive variation.

Here, we tested whether CPM-derived functional networks capture variability in sustained attention performance across participants in three independent datasets. All three datasets included fMRI tasks that required sustained attention to stimuli presented either in the visual or auditory domain. However, tasks differed across datasets in several ways, including frequency of responding, selection demands, and inhibitory control. Therefore, successful prediction of performance in these datasets suggests that functional networks successfully capture a signal of

sustained attention which is general across all three task contexts, rather than a distinct process idiosyncratic to a subset.

We first validated that a network of sustained attention previously defined using a CPM approach, the saCPM, generalizes to predict sustained attention performance in these datasets. Previous work has demonstrated that the saCPM captures patterns of connectivity related to attention by predicting out-of-sample performance on multiple attention tasks (Fountain-Zaragoza et al., 2019; Kardan et al., 2022; Rosenberg et al., 2018, 2020; Yoo et al., 2022) as well as attention-deficit/hyperactivity disorder symptomatology (Rosenberg et al., 2016) and variability in narrative engagement within individuals (Song et al., 2021). A previous study also found that network strength in the saCPM during rest predicted performance on an auditory sustained attention task (Wu et al., 2020). Our results show that the saCPM generalizes to predict sustained attention across perceptual modalities from task connectivity, further demonstrating that it captures domain-general signatures of attentional ability. Whereas previous work speculated that selective generalization of the saCPM to audiovisual movie engagement, but not audio-only story engagement, was due to modality specificity of the model (Song et al., 2021), our results find no modality bias when predicting individual differences in visual and auditory sustained attention. Rather, the differences in prediction observed in previous work may instead reflect other differences between stimuli, for example, in the overall engagement with the narratives.

We further show that models trained on sustained attention tasks performed in separate visual and auditory modalities generalize to predict sustained attention performance both in external datasets and when tasks were performed in a different perceptual modality than in training. This suggests that CPM identifies edges that capture variability in sustained attention performance that is not specific to the perceptual modality of the task. These results support previous findings

that the ability to sustain attention to visual and auditory information relies to some extent on shared neural mechanisms (Corriveau et al., 2024; Seli et al., 2012; Terashima et al., 2021). We used a CPM approach to identify a subset of edges that significantly predicted both auditory and visual sustained attention performance across individuals in a dataset. This subset of edges predicted both visual and auditory sustained attention performance in independent datasets. Therefore, this overlapping network of edges provides one mechanism that may support a modality-general ability to sustain attention over time.

Importantly, we show that successful generalization across modalities is not simply a mathematical inevitability due to correlations between sustained attention performance across modalities. While performance was reliable across participants regardless of task modality, generalization across modality persisted after controlling for performance in the other task modality during both model training and model testing. Therefore, predictive edges identified by CPM were able to capture relevant variance in sustained attention beyond consistency in performance.

Intriguingly, we observed significant prediction both from overlapping visual and auditory edges as well as modality-specific edges identified using a CPM approach. Therefore, feature reliability, or the identification of the same model features across training sets, was not necessary for successful generalization. These results highlight a distinction between model feature reliability and the ability to predict behavioral phenotypes in an external sample. Previous work has noted this difference, demonstrating that predictive accuracy is not necessarily a result of reliable features (Kragel et al., 2021; Noble et al., 2017; Tian & Zalesky, 2021, although see Chen et al., 2023). Researchers have suggested that a lack of reliability may be a function of the scale at which features are identified, leading to high numbers of model features (Srivastava et al.,

2022; Tian & Zalesky, 2021). Here, model features were identified from whole-brain patterns of functional connectivity, consisting of > 35,000 pairwise connections between regions. Therefore, it is difficult to determine whether the failure of an edge to be significantly related to performance in both visual and auditory networks is the result of the modality specificity of the edges or a result of the relatively small scale at which features were identified. As a result, edges identified only in one training set may capture modality-general sustained attention, leading to the significant prediction across modalities observed in the current study.

Individual edge contributions to auditory and visual networks from canonical functional networks varied. We found that connectivity within the default mode network was represented in the visual high-attention and overlapping high-attention networks but did not significantly contribute to auditory predictive networks. Much previous work has related relative increases in default mode network activation with in-the-zone attentional states (Esterman et al., 2013, 2014; Fortenbaugh et al., 2018; Jones et al., 2024; Kucyi et al., 2016, 2017; Song et al., 2023), although changes in activity are not functionally equivalent to changes in connectivity. Past work has shown links between greater within-default mode network connectivity and higher attention (Gordon et al., 2014; Kucyi & Davis, 2014). Further, attention-related disorders are characterized by decreased connectivity within the default mode network (Castellanos et al., 2008; Fair et al., 2010). However, other work has found an inverse or no relationship (Esterman et al., 2013; Kucyi et al., 2017; Mittner et al., 2014), suggesting associations of within-network connectivity of the default mode network with sustained attention are complex. The current findings suggest that stronger within-default mode network connections are associated with higher modality-general sustained attention performance.

We observed a large contribution of within-network edges from the subcortical-cerebellar and motor networks to visual and auditory low-attention networks, as well as the overlapping low-attention network. This is in line with previous work which has implicated greater within-subcortical-cerebellar connectivity in lower sustained attention performance (Fong et al., 2019; Jones et al., 2024; Rosenberg et al., 2016). Increased within-motor connectivity has similarly been related to poor sustained attention in adolescents, whereas connections between motor and visual regions are increased with better sustained attention (O'Halloran et al., 2018). We observed a similar pattern of results, with connectivity between motor and visual II networks contributing to visual and auditory high-attention networks. We do not see a significant contribution of motor to visual II connectivity to the overlapping high-attention network, suggesting the individual edges may differ between visual and auditory networks. Since the gradCPT used to train networks in the current study requires a motor (button press) response, it is possible that connections within and between the motor network are more strongly represented in these networks than would be expected if a different sustained attention task were used for network training. Future work may seek to test the extent to which task demands influence network architecture.

We should also note a few limitations of the current study. First, our analyses utilized a CPM approach, which sought to identify connections between brain regions whose strength captured variability in modality unique or modality general sustained attention. However, it is likely that functional relationships in the brain, beyond those at the edge level, may differ between task modality. While outside the scope of the current manuscript, future work may aim to more fully characterize functional differences between task, for example, at the level of graph-theoretic differences between whole-brain connectivity patterns. An additional limitation is the precision of

the current predicted sustained attention performance values. Significant correlations between predicted and observed sustained attention performance suggest that our sustained attention networks capture reliable differences in performance and are therefore useful in understanding neural mechanisms involved in sustained attention. However, the current models leave much variance unexplained, which may result from a number of individual, task, and dataset differences. Work aimed at precise predictions of sustained attention performance may choose to include additional variables in predictive models that better-capture this remaining variability.

While the current analyses focused on the generalization of sustained attention networks, a similar question could be asked of predictive networks trained on any cognitive process that can be performed in separate perceptual modalities. For example, it is an open question whether a network trained to predict visual recognition memory across participants would also generalize to predict auditory recognition memory, which is reliably worse (Cohen et al., 2009). Future work testing the validity of brain-based models of cognition should aim to test model generalizability across perceptual modalities to evaluate the extent to which a cognitive process is fully captured by a given model.

Our results demonstrate that functional connectivity-based networks of sustained attention are not specific to the perceptual modality of training, suggesting that these networks capture domain- and modality-general aspects of attention. Both nonoverlapping and overlapping, modality-general edges predicted cross-modal sustained attention performance in independent datasets, thereby providing one mechanism by which modality-general sustained attention ability may be supported. These results highlight that the ability to sustain attention to information over time relies on distributed, modality-general connections in the brain and demonstrate the potential for highly generalizable predictive models constructed from functional connectivity features.

**Chapter 2 Bibliography**


Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., … Chun, M. M. (2020). Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. Journal of Cognitive Neuroscience, 32(2), 241–255. https://doi.org/10.1162/jocn_a _01487, PubMed: 31659926

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., … Milham, M. P. (2008). Cingulate-precuneus interactions: A new locus of dysfunction in adult attentiondeficit/ hyperactivity disorder. Biological Psychiatry, 63(3), 332–337. https://doi.org/10.1016/j.biopsych.2007.06.025, PubMed: 17888409

Chen, J., Ooi, L. Q. R., Tan, T. W. K., Zhang, S., Li, J., Asplund, C. L., … Yeo, B. T. T. (2023). Relationship between prediction accuracy and feature importance reliability: An empirical and theoretical study. NeuroImage, 274, 120115. https://doi.org/10 .1016/j.neuroimage.2023.120115, PubMed: 37088322

Chidharom, M., Krieg, J., & Bonnefond, A. (2021). Impaired frontal midline theta during periods of high reaction time variability in schizophrenia. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6(4), 429–438. https://doi.org/10.1016/j .bpsc.2020.10.005, PubMed: 33431347

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proceedings of the National Academy of Sciences, 106(14), 6008–6010. https://doi.org/10.1073/pnas.0811884106, PubMed: 19307569

Corriveau, A., James, A. R., Jr., deBettencourt, M. T., & Rosenberg, M. D. (2024). Sustained attentional state is a floodlight not a spotlight. PsyArxiv. https://doi.org/10.31234/osf.io/k9cnm

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, 29(3), 162–173. https://doi.org/10.1006 /cbmr.1996.0014, PubMed: 8812068

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, PubMed: 22941724

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, PubMed: 24478354

Fair, D. A., Posner, J., Nagel, B. J., Bathula, D., Dias, T. G. C., Mills, K. L., … Nigg, J. T. (2010). Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. Biological Psychiatry, 68(12), 1084–1091. https://doi.org/10.1016/j .biopsych.2010.07.003, PubMed: 20728873

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. Nature Neuroscience, 18(11), 1664–1671. https://doi .org/10.1038/nn.4135, PubMed: 26457551

Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., … Chun, M. M. (2019). Dynamic functional connectivity during task performance and rest predicts individual differences in attention across studies. NeuroImage, 188, 14–25. https:// doi.org/10.1016/j.neuroimage.2018.11.057, PubMed: 30521950

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, PubMed: 29307606

Fountain-Zaragoza, S., Samimy, S., Rosenberg, M. D., & Prakash, R. S. (2019). Connectome-based models predict attentional control in aging adults. NeuroImage, 186,1–13. https://doi.org/10 .1016/j.neuroimage.2018.10.074, PubMed: 30394324

Gao, M., Wong, C. H. Y., Huang, H., Shao, R., Huang, R., Chan, C. C. H., & Lee, T. M. C. (2020). Connectome-based models can predict processing speed in older adults. NeuroImage, 223, 117290. https:// doi.org/10.1016/j.neuroimage.2020.117290, PubMed: 32871259

García-Pérez, M. A. (2023). Use and misuse of corrections for multiple testing. Methods in Psychology, 8, 100120. https://doi.org /10.1016/j.metip.2023.100120

Gordon, E. M., Breeden, A. L., Bean, S. E., & Vaidya, C. J. (2014). Working memory-related changes in functional connectivity persist beyond task disengagement. Human Brain Mapping, 35(3), 1004–1017. https://doi.org/10.1002/hbm.22230, PubMed: 23281202

Jones, H. M., Yoo, K., Chun, M. M., & Rosenberg, M. D. (2024). Edge-based general linear models capture moment-to-moment fluctuations in attention. Journal of Neuroscience, 344(14), e1543232024. https://doi.org/10.1523/JNEUROSCI.1543-23 .2024, PubMed: 38316565

Karamacoska, D., Barry, R. J., & Steiner, G. Z. (2018). Electrophysiological underpinnings of response variability in the Go/NoGo task. International Journal of Psychophysiology, 134, 159–167. https://doi .org/10.1016/j.ijpsycho.2018.09.008, PubMed: 30266622

Kardan, O., Stier, A. J., Cardenas-Iniguez, C., Schertz, K. E., Pruin, J. C., Deng, Y., … Rosenberg, M. D. (2022). Differences in the functional brain architecture of sustained attention and working memory in youth and adults. PLoS Biology, 20(12), e3001938. https://doi.org/10.1371/journal.pbio.3001938, PubMed: 36542658

Katayama, J., & Polich, J. (1999). Auditory and visual P300 topography from a 3 stimulus paradigm. Clinical Neurophysiology, 110(3), 463–468. https://doi.org/10.1016/S1388-2457(98)00035-2, PubMed: 10363770

Kim, H. (2014). Involvement of the dorsal and ventral attention networks in oddball stimulus processing: A meta-analysis. Human Brain Mapping, 35(5), 2265–2284. https://doi.org/10.1002/hbm .22326, PubMed: 23900833

Kondo, H. M., Terashima, H., Ezaki, T., Kochiyama, T., Kihara, K., & Kawahara, J. I. (2022). Dynamic transitions between brain states predict auditory attentional fluctuations. Frontiers in Neuroscience, 16, 816735. https://doi.org/10.3389/fnins.2022 .816735, PubMed: 35368290

Kondo, H. M., Terashima, H., Kihara, K., Kochiyama, T., Shimada, Y., & K aw a h a r a, J. I. ( 2 0 2 3). Prefrontal GABA and glutamate-glutamine levels affect sustained attention. Cerebral Cortex, 33, 10441–10452. https://doi.org/10.1093/cercor /bhad294, PubMed: 37562851

Kragel, P. A., Han, X., Kraynak, T. E., Gianaros, P. J., & Wager, T. D. (2021). Functional MRI can be highly reliable, but it depends on what you measure: A commentary on Elliott et al. (2020). Psychological Science, 32(4), 622–626. https://doi.org/10.1177 /0956797621989730, PubMed: 33685310

Kucyi, A., & Davis, K. D. (2014). Dynamic functional connectivity of the default mode network tracks daydreaming. NeuroImage, 100, 471–480. https://doi.org/10.1016/j.neuroimage.2014.06 .044, PubMed: 24973603

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 of the United States of America, 113(48), 13899–13904. https://doi.org/10.1073/pnas.1611743113, PubMed: 27856733

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, PubMed: 26874182

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. Psychological Bulletin, 139(4), 870–900. https:// doi.org/10.1037/a0030694, PubMed: 23163491

Linden, D. E. J., Prvulovic, D., Formisano, E., Völlinger, M., Zanella, F. E., Goebel, R., & Dierks, T. (1999). The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. Cerebral Cortex, 9(8), 815–823. https://doi.org/10.1093/cercor/9.8.815, PubMed: 10601000

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1(1), 6–21. https://doi.org/10.1080/17470214808416738

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. Journal of Neuroscience, 34(49), 16286–16295. https://doi.org/10.1523 /JNEUROSCI.2062-14.2014, PubMed: 25471568

Noble, S., Spann, M. N., Tokoglu, F., Shen, X., Constable, R. T., & Scheinost, D. (2017). Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cerebral Cortex, 27(11), 5415–5429. https://doi.org/10 .1093/cercor/bhx230, PubMed: 28968754

O'Halloran, L., Cao, Z., Ruddy, K., Jollans, L., Albaugh, M. D., Aleni, A., … Whelan, R. (2018). Neural circuitry underlying sustained attention in healthy adolescents and in ADHD symptomatology. NeuroImage, 169, 395–406. https://doi.org/10.1016/j.neuroimage .2017.12.030, PubMed: 29274748

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: A review. JAMA Psychiatry, 77(5), 534–540. https://doi.org/10.1001 /jamapsychiatry.2019.3671, PubMed: 31774490

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, PubMed: 9204482

Rosenberg, M. D., & Finn, E. S. (2022). How to establish robust brain–behavior relationships without thousands of individuals. Nature Neuroscience, 25(7), 835–837. https://doi.org/10.1038/s41593-022-01110-9, PubMed: 35710985

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. Nature Neuroscience, 19(1), 165–171. https://doi.org/10.1038/nn .4179, PubMed: 26595653

Rosenberg, M. D., Zhang, S., Hsu, W.-T., Scheinost, D., Finn, E. S., Shen, X., … Chun, M. M. (2016). Methylphenidate modulates functional network connectivity to enhance attention. Journal of Neuroscience, 36(37), 9547–9557. https://doi.org/10.1523 /JNEUROSCI.1746-16.2016, PubMed: 27629707

Rosenberg, M. D., Hsu, W.-T., Scheinost, D., Todd Constable, R., & Chun, M. M. (2018). Connectome-based models predict separable components of attention in novel individuals. Journal of Cognitive Neuroscience, 30(2), 160–173. https://doi.org/10.1162 /jocn_a_01197, PubMed: 29040013

Rosenberg, M. D., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., … Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. Proceedings of the National Academy of Sciences of the United States of America, 117(7), 3797–3807. https://doi.org/10 .1073/pnas.1912226117, PubMed: 32019892

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. Attention, Perception, & Psychophysics, 75(3), 426–439. https://doi.org/10.3758 /s13414-012-0413-x, PubMed: 23299180

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., … Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. NeuroImage, 193, 35–45. https://doi.org/10.1016/j.neuroimage.2019 .02.057, PubMed: 30831310

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66(1), 44–50. https://doi.org/10.1037/a0025111, PubMed: 21910522

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome based predictive modeling to predict individual behavior from brain connectivity. Nature Protocols, 12(3), 506–518. https://doi .org/10.1038/nprot.2016.178, PubMed: 28182017

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. NeuroImage, 82, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081, PubMed: 23747961

Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., … Rorden, C. (2010). Spatial attention evokes similar activation patterns for visual and auditory stimuli. Journal of Cognitive Neuroscience, 22(2), 347–361. https://doi.org/10.1162/jocn.2009 .21241, PubMed: 19400684

Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. Proceedings of the National Academy of Sciences, 118(33), e2021905118. https://doi.org/10.1073/pnas .2021905118, PubMed: 34385312

Song, H., Shim, W. M., & Rosenberg, M. D. (2023). Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. eLife, 12, e85487. https://doi .org/10.7554/eLife.85487, PubMed: 37395724

Srivastava, P., Fotiadis, P., Parkes, L., & Bassett, D. S. (2022). The expanding horizons of network neuroscience: From description to prediction and control. NeuroImage, 258, 119250. https://doi.org/10.1016/j.neuroimage.2022.119250, PubMed: 35659996

Stevens, A. A., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. Magnetic Resonance Imaging, 18(5), 495–502. https://doi.org/10.1016 /S0730-725X(00)00128-4, PubMed: 10913710

Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., & Epstein, J. N. (2012). Reaction time variability in ADHD: A review. Neurotherapeutics, 9(3), 500–508. https://doi.org/10 .1007/s13311-012-0138-5, PubMed: 22930417

Terashima, H., Kihara, K., Kawahara, J. I., & Kondo, H. M. (2021). Common principles underlie the fluctuation of auditory and visual sustained attention. Quarterly Journal of Experimental Psychology, 74 (4), 705 –715. https://doi.org/10.1177 /1747021820972255, PubMed: 33103992

Tian, Y., & Zalesky, A. (2021). Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? NeuroImage, 245, 118648. https://doi.org/10.1016/j.neuroimage .2021.118648, PubMed: 34673248

Walz, J. M., Goldman, R. I., Carapezza, M., Muraskin, J., Brown, T. R., & Sajda, P. (2013). Simultaneous EEG-fMRI Reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. Journal of Neuroscience, 33(49), 19212–19222. https://doi.org/10.1523/JNEUROSCI.2649-13 .2013, PubMed: 24305817

Wu, E. X. W., Liaw, G. J., Goh, R. Z., Chia, T. T. Y., Chee, A. M. J., Obana, T., … Asplund, C. L. (2020). Overlapping attentional networks yield divergent behavioral predictions across tasks: Neuromarkers for diffuse and focused attention? NeuroImage, 209, 116535. https://doi.org/10.1016/j.neuroimage.2020.116535, PubMed: 31940476

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3485–3492. https://doi.org/10 .1109/CVPR.2010.5539970

Yoo, K., Rosenberg, M. D., Kwon, Y. H., Lin, Q., Avery, E. W., Sheinost, D., … Chun, M. M. (2022). A brain-based general measure of attention. Nature Human Behaviour, 6(6), 782–795. https://doi.org/10.1038/s41562-022-01301-1, PubMed: 35241793

# Chapter 3: Common brain network dynamics capture attention fluctuations in tasks and movies

Large-scale functional brain networks are constantly in flux. One ongoing goal of cognitive neuroscience is to relate dynamic neural fluctuations to changes in internal states and behaviors. To date, identification of dynamic functional networks has been limited in its temporal resolution, requiring calculation of functional connectivity over time windows. Recent methods, however, enable the exploration of functional network reconfiguration at individual points in time, removing this constraint (Faskowitz et al., 2020; Zamani Esfahlani et al., 2020). In the current study, we exploit time-resolved network analyses to identify functional networks that track fluctuations in sustained attention, which occur over a matter of seconds (Rosenberg & Song, 2021).

While sustained attention is ubiquitous throughout daily life, situations requiring continuous focus may differ in many ways. What we attend, the perceptual modality in which information is received, and the salience of information may differ between contexts, among other factors. For example, while reading an email and conversing with a friend both require the ability to attend and understand a stream of words, these activities require processing of different perceptual modalities and may differ dramatically in their inherent ability to exogenously capture attention. Despite these discrepancies, work examining individual differences shows that sustained attention performance is consistent across task contexts (Unsworth et al., 2021) and perceptual modalities (Corriveau et al., 2025b; Seli et al., 2011; Terashima et al., 2021), supporting a core sustained attentional construct. Therefore, generalizable brain-based models of sustained attention must capture context- and modality-agnostic features of sustained attention.

Interactions between brain regions measured using functional magnetic resonance imaging (fMRI) have repeatedly been linked to sustained attention. Functional MRI connectivity-based

networks capture differences in sustained attention ability across individuals (Corriveau et al., 2022; Rosenberg et al., 2016) and are highly generalizable, predicting individual differences across different attention tasks (Yoo et al., 2022), perceptual modalities (Corriveau et al., 2025a), and age groups (Kardan et al., 2022). However, sustained attention not only varies across individuals but also within an individual on the order of seconds (Castellanos et al.., 2005; Rosenberg et al., 2025; Terashima et al., 2021). Work investigating fluctuations over time using block-averaged connectivity (Rosenberg et al., 2020; Kardan et al., 2022), sliding-window connectivity (Song et al., 2021), or psychophysiological interaction analysis (Kucyi et al., 2017) has shown that functional connections reconfigure with changes in attentional state. These previous methods, however, involve averaging or smoothing across time windows which may obscure high-frequency fluctuations in attention.

A recently-developed method, termed edge time series, computes functional interactions between brain regions at a single time point, eliminating the need to average or smooth functional connectivity data over time (Faskowitz et al., 2020; Zamani Esfahlani et al., 2020). In fMRI, edge time series reflect the element-wise product of z-scored univariate activity time courses. Because this calculation can be performed on a single time point, edge time series enable the identification of functional connections that correspond with high-frequency changes in cognition and behavior. Using methodology akin to univariate general linear models (GLMs), edge-based GLMs can identify networks of edges whose interaction, or co-fluctuation, tracks differences in sustained attentional state from trial-to-trial (Jones et al., 2024). Importantly, while mathematically equivalent to the interaction between univariate time courses, edge time series provide explanatory power above and beyond univariate activity (Merritt et al, 2024). Therefore, edge time series

enable the identification of neural substrates of dynamic cognition and attention at a higher resolution than previously thought possible.

An additional barrier to the identification of functional networks related to cognition and attention is their generalization. Valid functional networks should capture variation in a new dataset that may differ in many ways from data used to train the networks. With the recent development of edge-based GLMs, relatively little work has tested their ability to generalize across task contexts. Here, we explore the robustness of edge-based methods for identifying functional networks related to sustained attention performance in a controlled task and use these networks to predict attentional fluctuations in a different, naturalistic paradigm. This framework provides a case study for the validity of novel edge-based methods as predictive tools.

In the current study, we leverage edge-based GLMs to demonstrate that high-frequency fluctuations in sustained attention to visual and auditory information involve not only similar univariate brain activity, but also similar functional network reconfiguration. Importantly, we find that univariate activity and functional networks provide unique insights into neural mechanisms of sustained attention, demonstrating utility of the edge-based approach. Finally, we show that edge-based networks generalize between distinct task contexts. Results highlight the robustness of time-resolved functional networks for the study of sustained attention fluctuations across unique perceptual modalities and tasks.

## Methods

Data was collected at the MRI Research Center at the University of Chicago. Participants (N=60) completed at least one session of a two-session fMRI study. Functional MRI data were part of a large data collection effort and sample size was selected to provide high power for

multiple planned analyses. Sessions were completed approximately one week apart (mean time between sessions=10.88 days, SD=9.87 days). During both sessions, participants viewed or listened to naturalistic narrative stimuli, performed a sustained attention task, and completed an annotated rest task in the fMRI scanner. Scan order for a single session was as follows: single-modality (auditory-only or visual-only) naturalistic stimulus, annotated rest, audio-visual naturalistic stimulus, annotated rest, sustained attention task. Annotated rest task data are not analyzed here. Functional MRI data from the sustained attention and annotated rest tasks have been reported previously using analyses independent of the current manuscript (Corriveau et al., 2025a; Ke et al., 2025).

Functional MRI data were collected on a 3T Philips Ingenia scanner. Voxels were 2.526mm x 2.526mm x 3mm. Volumes were collected using a multiband sequence with a repetition time of 1 second. Three volumes were removed from the start of each scan. All study procedures were approved by the Social and Behavioral Science Institutional Review Board at the University of Chicago.

Following the fMRI scan, participants completed a series of behavioral tasks. Participants watched or listened again to the naturalistic narrative stimuli presented during scanning and provided a continuous engagement rating on a sliding scale (described in detail below). After the narrative, participants also provided one value on the sliding scale rating how engaging they found the stimulus overall. Prior to this second presentation of narrative stimuli, participants also completed memory tasks for the naturalistic stimuli as well as images and sounds presented during the sustained attention task. Memory data and summary engagement scores are not analyzed here.

*Audio-visual continuous performance task*

Sustained attention performance was tested in both sessions using a 10-minute audio-visual continuous performance task (avCPT; Corriveau et al., 2025b). During the avCPT, participants were presented a stream of trial-unique images and sounds presented simultaneously. Images were presented continuously for 1.2 seconds each whereas sounds were presented for 1 second with a 200 ms inter-trial interval to allow participants to distinguish individual sounds. Participants made a category-frequency judgement on either the images or the sounds such that both modalities served as the task-relevant modality across the two scanning sessions. Order was counterbalanced across participants. Participants were instructed to press a button when a stimulus from the task-relevant modality belonged to the frequent category (indoor/outdoor images, natural/manmade sounds) and withhold a button press to infrequent stimuli.

In rare cases, participants may have responded slowly to a frequent-category stimulus after the onset of the following trial. In this situation, it would appear that the participant made two key presses during the following trial when the first key press was intended as a response to the preceding trial. To account for this, we performed response reassignment if the following conditions were met: 1. A participant responded more than once during a frequent-category trial, 2. the first response was made within 100ms of the trial onset, and 3. no response was made to the previous frequent-category trial. In these cases, the first key press would be attributed to the preceding trial. Response reassignment was rare in both visual-relevant (mean reassignments = 0.548, SD=0.861) and auditory-relevant (mean reassignments=3.81, SD=3.46), affecting fewer than 0.8% of trials.

Sustained attention performance was behaviorally quantified in two ways. Accuracy of responses to infrequent stimuli served as momentary probes of sustained attentional state, with

correctly withheld responses indicating high attention and failure to withhold a response indicating a sustained attentional lapse. Additionally, sustained attentional state was evaluated continuously using variance time course (VTC) which identifies periods of consistent, in-the-zone responding and erratic, out-of-the zone responding (Esterman et al., 2013). The VTC quantifies deviation response times, such that low VTC values indicate periods of stable responding and therefore better sustained attention, whereas high VTC values are indicative of worse sustained attention. The VTC was computed using correct responses to relevant, frequent-category trials, as these were the trials on which a response should have been made. Response time courses were linearly detrended to account for speeding over time. Response times were z-scored, resulting in a time course in which the mean is 0. Next, all values in the time course were converted to absolute values. Finally, missing trial responses were linearly interpolated using responses from the two adjacent trials. The resulting vector reflects a time course of the extent to which each response deviates from the mean.

Logistic mixed-effect models were used to test whether variability in response times predicted successful performance on infrequent trials during the avCPT. Models included a fixed effect of preceding VTC, calculated as the average VTC during the three trials preceding an infrequent probe trial, as well as a subject-level random intercept. Models were fit for auditory-relevant and visual-relevant runs separately.

*Naturalistic narrative stimuli*

Participants were presented four naturalistic narrative stimuli across both scan sessions—an auditory-only podcast, a silent movie, and two audio-visual movies. The podcast ("*Pie*," 16 min 34 s) consisted of a baker describing the process of baking a berry pie. The silent movie ("*Croissant*," 16 min 59 s) featured clips from a baker in the process of making strawberry and

cream-filled croissants. One audio-visual movie ("*Cake*," 15 min 40 s) featured the baking of a tiered chocolate cherry cake with narration. The final audio-visual stimulus (14 min 49 s) was a clip from the movie *North by Northwest* including dialogue between two main characters and a suspenseful airplane chase scene.

Following both scans, participants watched or listened to each stimulus a second time and performed an engagement rating task. Task instructions were to continuously indicate their subjective level of engagement using arrow keys to adjust the position of a slider bar on a scale from "Not engaging at all" to "Completely engaging." Participants could indicate a large change in engagement by pressing the shift key and an arrow key simultaneously. The sliding scale contained 25 steps and the bar always initialized on step 13. Single key presses moved the bar one step and large jumps (shift+arrow key) moved the bar 5 steps.

*fMRI preprocessing*

Functional MRI data were preprocessed using Analysis of Functional Neuroimages (AFNI version 19.0). The first 3 TRs of each functional run were excluded from analysis. Despiking, slice-time correction, and motion-correction was applied to the data. We regressed covariates of no interest from the data, including a 24-parameter head motion model (6 motion parameters, 6 temporal derivatives, and their squares) and mean signal from subject-specific eroded white matter and ventricle masks and the whole brain. Additional motion censoring was performed on avCPT runs: volumes in which more than 10% of voxels were outliers or the Euclidean distance of the head motion parameter derivatives were greater than 0.25 mm were removed. We then registered the functional images to participants' skull-stripped MPRAGE anatomical images with linear

transformation and then normalized to the Montreal Neurological Institute (MNI) space with nonlinear warping.

Functional data for all runs were parcellated into 400 cortical and 32 subcortical nodes using the Schaefer 400-node atlas (Schaefer et al., 2018) and the Melbourne Subcortical Atlas Scale II (Tian et al., 2020), respectively. This resulted in 432 time series for each run.

*Data exclusion*

Functional MRI data from avCPT runs were excluded if in-scanner head motion exceeded any of the following criteria: maximum framewise displacement after censoring > 4mm, average framewise displacement after censoring > 0.15mm, or greater than 50% of volumes censored. Of the 57 participants who completed an auditory-relevant run of the avCPT, 10 were removed for head motion. For visual-relevant avCPT runs, fMRI data was available for 56 participants, 9 of whom were removed for head motion. Visual-relevant data was excluded from one additional participant who was missing data from 134 ROIs (~31% of the brain). Additionally, we did not analyze fMRI data from participants who performed more than 2.5 standard deviations below mean performance (*A'*) across all auditory and visual runs of the avCPT task. One participant's auditory-relevant run and one participant's visual-relevant run were removed for low performance. This left a total of 46 participants with auditory-relevant and 45 participants with visual-relevant fMRI data. Of these, 37 participants are included in both analyses.

Naturalistic fMRI runs were used for model testing. Therefore, to maximize the number of possible testing datasets and because excessive motion could only impair model generalization, we did not perform exclusions for head motion during naturalistic runs. Neural data from one participant watching *North by Northwest* and one participant listening to *Pie* were excluded

because data was missing >100 nodes. Final analyses of fMRI naturalistic data were performed on the following sample sizes: *Croissant* N=57, *Pie* N=55, *Cake* N=57, *North by Northwest* N=58.

We excluded behavioral engagement rating data from participants who reported a change in engagement less than once per minute on average, suggesting task noncompliance. Three engagement ratings from *Pie*, three from *Croissant*, four from *Cake*, and three from *North by Northwest* were excluded due to infrequent responding. Mean engagement time courses were calculated from the following sample sizes: 54 from *Pie*, 55 from *Croissant*, 50 from *Cake*, and 54 from *North by Northwest*.

*Univariate activity related to sustained attention*

We conducted a general linear model analysis on BOLD activity to identify brain regions that track fluctuations in sustained attention. We fit two first-level models with sustained attention regressors of interest at the ROI level (Fortenbaugh et al., 2018; Jones et al., 2024). Design matrices were constructed using nilearn's *make_first_level_design_matrix* function. The first model included three attention-related regressors: incorrect presses to relevant, infrequent targets (lapses), correctly-withheld responses to relevant, infrequent targets (correct omissions), and incorrect failures to respond to relevant, frequent stimuli. Trial types were modelled with impulse regressors. Individual participants' time series were fit using the autoregressive (AR) model in nilearn. Additional impulse regressors were included in the model for each censored time point. Regressors were convolved with the SPM HRF model as well as its temporal derivative. To identify ROIs in which activity differed between high attention and lapses, we computed the contrast: correct omissions – lapses.

The second model sought to identify activity related to continuous fluctuations of sustained attention. This model included a parametric regressor of interest, the VTC, which provides a continuous measure of sustained attentional state. Additionally, a boxcar regressor of constant duration was used to model individual trials and a boxcar regressor of variable duration, in which the duration reflected the linearly-detrended response time on each trial (Mumford et al., 2023), was included to account for activity that reflects response speed alone. Again, models were fit using nilearn's AR model and impulse regressors were included for censored time points. Regressors were convolved with the SPM HRF model and its temporal derivative. To identify regions whose activity fluctuated with sustained attention, we computed the contrast VTC over baseline.

To correct for family-wise error rates, max-T thresholding was performed for univariate analyses. T-statistics were converted to F-statistics for permutation testing. Observed F-statistics were compared to null distributions generated by permuting F-statistics across ROIs 10,000 times. Significant ROIs were those whose observed statistics were more extreme than the null distribution, $p<.05$.

*Edge co-fluctuation networks related to sustained attention*

Edge co-fluctuation time series quantify high-frequency changes in activity synchrony between pairs of brain regions (Faskowitz et al., 2020; Zamani-Esfahlani et al., 2020). Intuitively, edge time series "unravel" the Pearson correlation by quantifying two regions' simultaneous deflection from their mean activity level. Computing the average of an edge time series returns the Pearson correlation between two regions' activity time courses. For all pairs of brain regions, we calculated the edge time series as the element-wise product between z-scored BOLD activity time

courses. The resulting vector quantifies the extent to which, at each time point, the regions were activating (or deactivating) similarly relative to their mean.

Importantly, we can fit general linear models to edge time series to identify pairs of brain regions whose co-fluctuations track sustained attention performance (Jones et al., 2024). We fit the two models described above to edge co-fluctuation time series to identify edge networks involved in successful vs. lapsing sustained attentional states, as well as continuous attentional fluctuations using the variance time course.

Significant edges were those whose F-statistics were more extreme ($p<.05$) than a random null, created by sign-flipping the true edge values 10,000 times. To control for false positives, we identified edges that significantly tracked sustained attention performance across participants in both auditory-relevant and visual-relevant runs. This set of edges can be thought of as a modality-general edge network, as it tracked sustained attention regardless of modality in which stimuli were presented. This overlapping network also ensures robustness such that it includes edges whose prediction of sustained attention fluctuations internally replicate. While not performed in the current study, an alternative method for controlling the false positive rate in edge co-fluctuation networks within runs rather than across runs is to use the network based statistic, which controls for family-wise error rate by assessing the size of connected network components relative to a null distribution, analogous to cluster-based thresholding of activation results (Zalesky et al., 2010, see Discussion). Significance of overlap between runs was evaluated comparing the true number of edges that appeared in both runs to a null distribution created by randomly permuting the rows and columns of the significant edge matrices in both visual-relevant and auditory-relevant runs and counting the number of edges that were shared between runs, 10,000 times.

To evaluate the contribution of canonical networks to sustained attention, significant edges were grouped into 8 networks (7 canonical networks from Yeo et al., 2011 and 1 subcortical network from Tian et al., 2020). The significance of canonical network involvement was calculated by comparing the number of significant edges belonging to a canonical network to a null distribution calculated by permuting rows and columns of edge matrices, 10,000 times. Significant networks were those whose true edge membership was greater than the shuffled distribution, Bonferroni corrected for the number of within- and between-network comparisons made, $p<(.05/36)$.

*Theoretical edge networks predicted from univariate activity*

To examine the unique contribution of edge co-fluctuations to predictive power of sustained attention performance, over and above contributions of ROI activity alone, we compared edge networks that tracked attention fluctuations to theoretical edge maps predicted by univariate activity. Univariate-predicted edge maps were calculated as the signed square root of the product of t-statistics between all pairs of brain regions (Jones et al., 2024):

$$\hat{e}_{ij} = sign(t_i \cdot t_j) \cdot \sqrt{|t_i \cdot t_j|}$$

In this formula, $t_i$ and $t_j$ refer to t-statistics for regions $i$ and $j$ resulting from the univariate GLMs described previously. The resulting value $\hat{e}_{ij}$ is calculated for all pairs of brain regions and provides a theoretical edge value that should be expected based on the assumption that if activity in two regions is similarly related to a third variable of interest (in this case, sustained attention performance), their co-fluctuation time series should also be related to this third variable.

We calculated $\hat{e}_{ij}$ for all pairs of brain regions, resulting in a theoretical map of edges expected to be related to sustained attention based on univariate activity. Significant theoretical

edges were those whose F-statistics ($\hat{e}_{ij}^2$) were significant ($p<.05$). We then compared the map of theoretical edges identified to be significant in both visual-relevant and audio-relevant runs of the avCPT to the same map of observed edges. If the edges predicted in the full theoretical edge map largely overlap with the edges observed to track sustained attention performance, it would suggest that edge-based predictions are redundant with ROI-based predictions. However, if ROI activity-based predictions are largely unique from observed edge relationships, it would suggest that edges provide additional explanatory power above and beyond what is predicted by univariate activity.

To provide intuition as to when univariate and edge-predicted time series result in similar vs. diverging predictions, we also performed a simulation analysis. We simulated pairs of theoretical univariate time series that differed in the extent to which they were related to a theoretical third variable as well as the extent to which they were related to each other. We then observed the relationship between the edge time series for each pair of univariate time series and the theoretical third variable. Finally, we tested this relationship as various levels of noise were applied to the data. Full methods are included in the **Appendix**.

*Generalization of edge networks to predict narrative engagement*

To test whether edge networks defined during a controlled sustained attention task also track attentional fluctuations in more naturalistic contexts, we compared predicted changes in attentional state based on edge network strength to subjective reports of engagement while participants watched and listened to naturalistic narratives. Using a leave-one-subject-out approach, we identified modality-general edge networks—that is, edges that were related to sustained attention successes vs. lapses or VTC in both auditory- and visual-relevant task runs—for N-1 subjects. Then, we calculated TR-by-TR edge network strength in the left-out subject while

they viewed and/or listened to each of the four naturalistic narrative stimuli. Edge network strength was defined as the difference in average edge co-fluctuation strength between positive (those with positive T-values) and negative (those with negative T-values) edges. This edge network strength calculation is based on previous work (Rosenberg et al., 2016) but is adapted to return the network strength at each time point. The difference in positive and negative edge co-fluctuation strength is used as a summary metric because predictions from these networks are anticorrelated by nature of how these edges are identified. The resulting time course reflects time-varying network strength across an entire naturalistic run.

We then correlated the modality-general edge network strength time series with subjective ratings of engagement across an entire run. If edge networks generalize to capture subjective engagement, this correlation should be stronger than a null distribution, created by circle-shifting the true or reversed network strength time course 10,000 times. Circle shifting was limited such that the shift had to be greater than 8 time points from the true time course. For main analyses, edge network strength was calculated at the participant-level and compared to the mean engagement time course across all participants, as the group mean may reflect a less-noisy estimate of stimulus-driven engagement (Song et al., 2021). Mean engagement was computed by z-scoring engagement time courses within participants and averaging time courses across participants. We include results where each individual's edge network strength time course was correlated with their unique z-scored subjective engagement time course in the **Appendix**.

For comparison, we also tested whether univariate activity in regions related to sustained attention during both visual-relevant and auditory-relevant runs of the avCPT tracked changes in engagement during naturalistic narratives. Predicted univariate time courses were calculated as the difference between average activity in regions positively and negatively related to sustained

attention during the avCPT. We correlated these predicted time courses with the average engagement time course across participants. Results of this analysis are reported in the **Appendix**.

## Results

Performance on the avCPT, measured as *A',* was high during both visual-relevant (M=0.937, SD=0.045) and auditory-relevant runs (M=0.744, SD=0.116), demonstrating participant compliance. A logistic regression predicting avCPT performance with a main effect of run type and a participant-level random intercept revealed that performance was significantly higher during visual-relevant runs ($b_{run\ type}$=0.192*, SE*=0.019*, p*<.001*)* than auditory-relevant runs, suggesting increased difficulty in the auditory task. Similar results were observed in a larger behavioral sample (Corriveau et al., 2025b).

Response time variance, quantified using the VTC, significantly predicted success on infrequent trials during the avCPT. Specifically, infrequent probes in which participants correctly withheld a button press (correct omissions) were preceded by more stable responding, and therefore lower preceding VTC, in both visual-relevant (*b*=-0.141, SE=0.053, p=$7.69*10^{-3}$) and auditory-relevant sessions (*b*=-0.135, SE=0.046, p=$3.33*10^{-3}$). These findings confirm that the VTC tracks ongoing fluctuations in sustained attention performance, replicating previous work (Esterman et al., 2013).

### *Auditory and visual sustained attention fluctuations recruit shared activity*

To investigate univariate activity involved in sustained attention, we explored evoked activity for the following GLM regressors: correct omissions vs. lapses and the variance time

course. Thresholded univariate maps are visualized in **Figure 11**. Unthresholded maps are included in **Appendix 2 Figure 41**.

We first investigated activity differentially related to correct omissions vs. failures of sustained attention (lapses) on infrequent trials (**Figure 11A**). In visual-relevant runs, visual and temporal regions were involved in successful sustained attention performance, whereas areas in the prefrontal cortex, frontal operculum/insula, posterior cingulate cortex (PCC), and subcortex were more strongly activated during failures of sustained attention. This pattern of results largely replicates previous work (Fortenbaugh et al., 2018; Jones et al., 2024). During auditory-relevant runs, successful sustained attention performance was characterized by increased activity in temporal and right-hemisphere somatomotor regions, while increased left-hemisphere PCC activity was associated with lapses. Regions common to both tasks included bilateral temporal areas involved in sustained attentional successes and left PCC involved in lapses.

We also examined activity related to the variance time course (**Figure 11B**). Regions showing positive statistics reflect areas whose activity increased with greater response time variance, or worse sustained attentional states, while regions showing a negative relationship are those whose activity increased with reductions in response variance, indicating better sustained attentional states. In both visual-relevant and auditory-relevant tasks, regions in attention and salience networks, such as the frontal operculum/insula, temporal parietal junction (TPJ), frontal eye fields (FEF), and anterior cingulate cortex (ACC), showed increased activity under worse sustained attentional states. These results closely replicate previous work (Esterman 2013; Fortenbaugh et al., 2018; Jones et al., 2024).

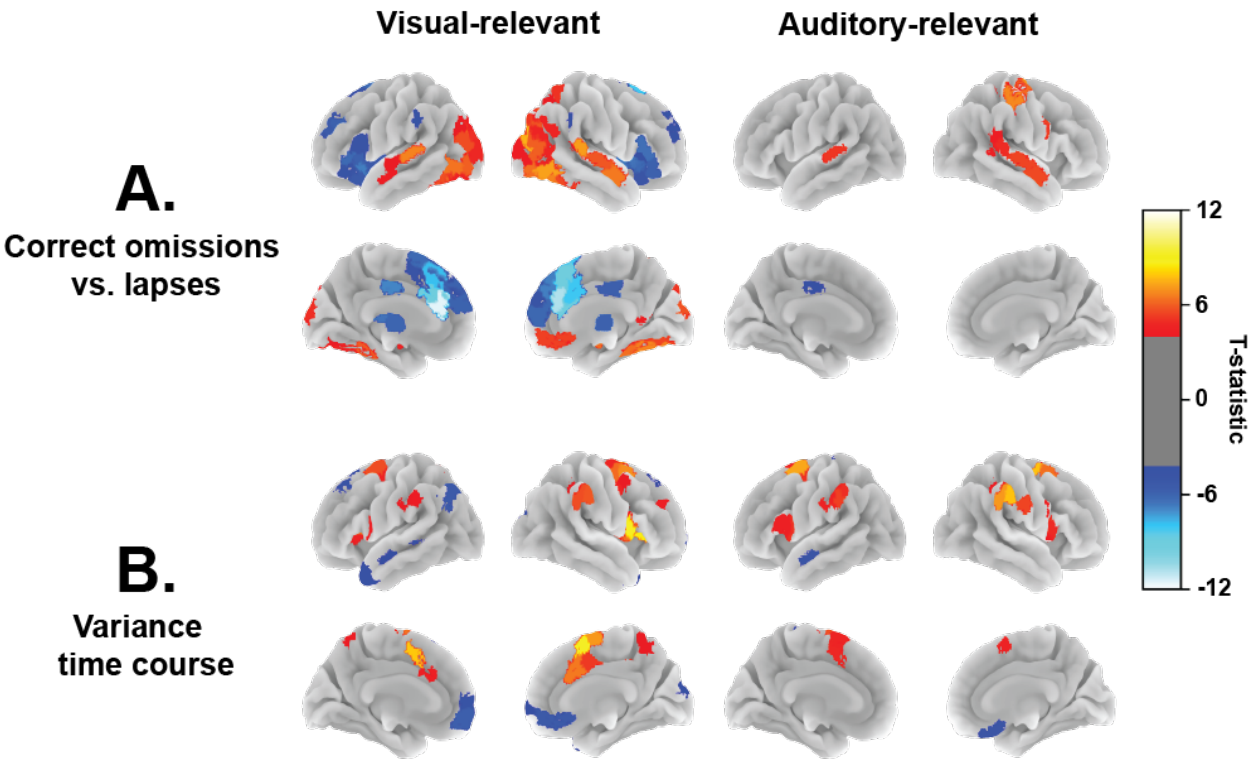



*Figure 11. Maps visualizing significant results from contrasts of interest.* (A) Difference maps for correct omissions vs lapses reveal regions in the temporal cortex and the PCC involved in both visual- and auditory-relevant sustained attention. Activity more strongly evoked by correct inhibition of responses to infrequent trials is visualized in warm colors, and activity related to failure to inhibit responses is visualized in cool colors. (B) Univariate activity associated with response time variance reveal regions in the dorsal and ventral attention networks that track sustained attention fluctuations. Regions in warm colors indicate areas whose activity increased with greater response time variance, reflecting worse sustained attentional states. Cool colors reflect regions whose activity increased with decreases in response variance, indicating better sustained attention. Colored regions reflect those surviving max-T thresholding, $p<.05$.

*Shared edge networks underlie auditory and visual sustained attention*

We next examined how network reconfiguration at the TR-by-TR time scale tracked changes in sustained attention. We fit our GLMs to edge co-fluctuation time series to identify edge networks related to our sustained attention contrasts of interest, i.e., correct omissions vs. lapses and the variance time course. Models were fit for visual-relevant and auditory-relevant fMRI runs separately. Resulting networks comprise the set of edges whose co-fluctuation time course significantly reflected changes in sustained attention during the avCPT.

Contrasting correct omissions vs. lapses revealed a set of 5,274 edges whose co-fluctuation was stronger during correct omissions and 6,654 edges whose co-fluctuation was stronger during attentional lapses during visual-relevant runs (**Figure 12A**). In auditory-relevant runs, 1,883 edges showed stronger co-fluctuation during correct omissions and 2,739 edges showed stronger co-fluctuation during lapses. Between visual-relevant and auditory-relevant runs, 160 edges positively tracking correct omissions and 271 edges positively tracking lapses were shared by both run types, a significant amount of overlap in both cases (both *p*s<.001; **Figure 12B**). In this modality-general overlap, edges connecting the ventral attention to default mode and limbic networks showed higher co-fluctuations during correct omissions. Conversely, edges within the default mode and ventral attention networks, as well as edges between the somatomotor network and subcortex showed stronger co-fluctuations during attentional lapses. Edge networks also shared edges in the unexpected direction, although, as predicted, this overlap was not significant. The network that positively tracked correct omissions in visual-relevant runs shared 91 edges with the network that tracked lapses during auditory-relevant runs, whereas the network that tracked correct omissions during auditory-relevant runs shared 69 edges with the visual-relevant lapse network (*p*s>.999).

When contrasting the variance time course against baseline in visual-relevant runs, 5,306 edges were positively related to the variance time course, i.e., showed higher co-fluctuation during worse sustained attentional states, whereas 5,507 edges showed a negative relationship with the variance time course. In auditory-relevant runs, 3,675 edges were positively related to the variance time course while 3,356 were negatively related. Both positive and negative edge networks showed significant overlap between the two sessions (790 overlapping positive edges, 403 overlapping negative edges; *p*s<.001). Overlapping edges positively associated with the variance time course, therefore showing stronger co-fluctuation during worse sustained attention, were found within the somatomotor and dorsal attention networks, between the dorsal attention and ventral attention networks, and between control networks and dorsal and ventral attention networks. Edges connecting the visual network to somatomotor, dorsal attention, and limbic networks were negatively associated with the variance time course, i.e., showed higher co-fluctuations during better sustained attention. Again, networks trained to differentially predict the variance time course between modalities did not significantly overlap. The network that positively tracked variance time course fluctuations during visual-relevant runs shared 58 edges with the negative network during auditory-relevant runs, while the negative network during visual-relevant runs shared 57 edges with the positive network during auditory-relevant runs (*p*s>.999). Results confirm that edge-based GLMs identify modality-general networks that reconfigure with fluctuations in sustained attention.

Interestingly, edge networks related to better sustained attention, that is, positive networks related to correct omissions vs. lapses and negative networks related to the variance time course, significantly overlapped in visual-relevant runs (1450 overlapping edges, *p*<.001) but not auditory-relevant runs (84 overlapping edges, *p*>.999). Similarly, networks related to worse sustained attention as measured by correct omissions vs. lapses and the variance time course were

significantly overlapping in visual-relevant runs (1932 overlapping edges, $p<.001$) but not auditory-relevant runs (196 overlapping edges, $p=.835$). Edges related to sustained attention across all networks were not significant (0 edges related to better sustained attention, 3 edges related to worse sustained attention, $ps>.999$), likely due to limited overlap between auditory-relevant runs.

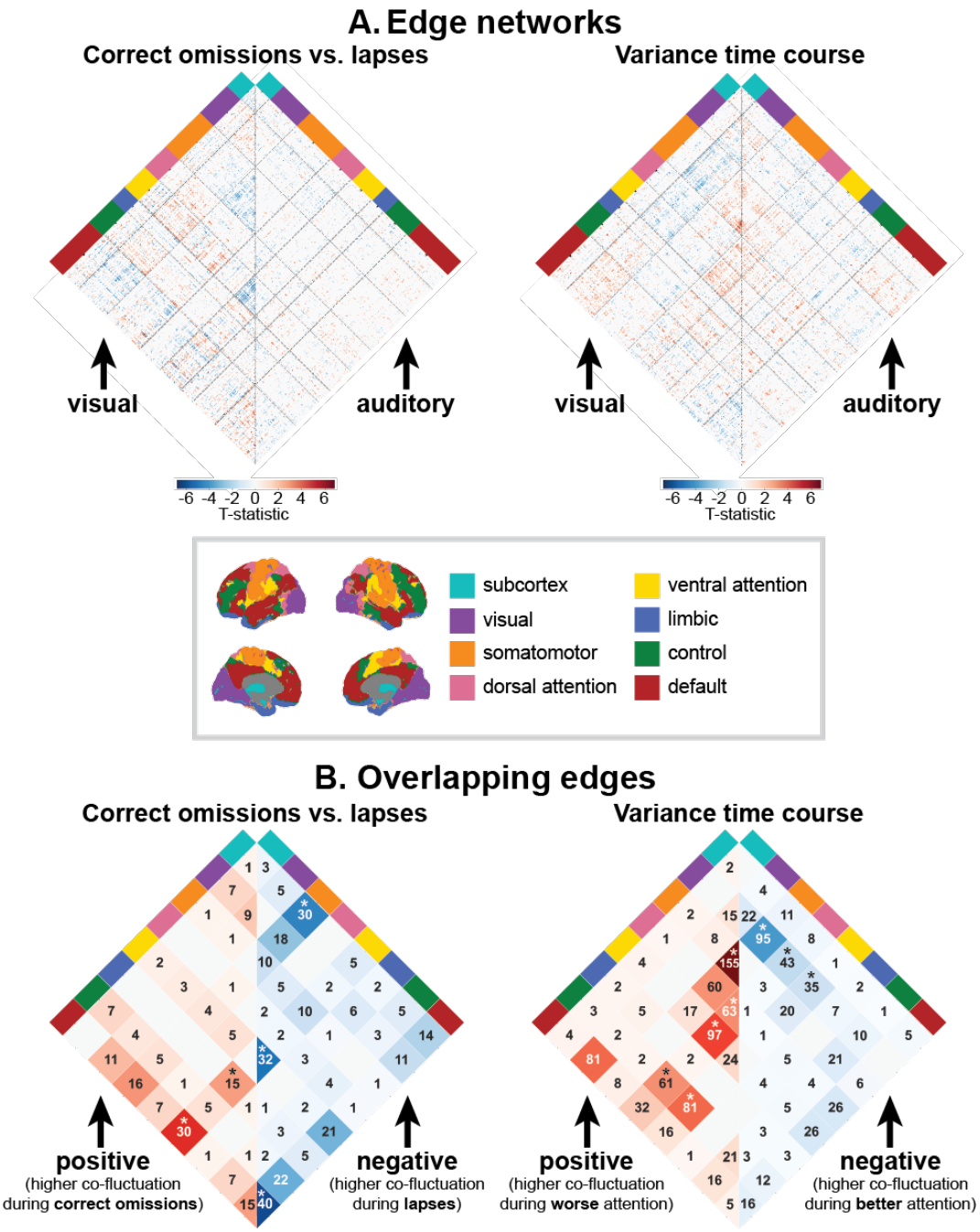



*Figure 12. Canonical network contributions to edge networks.* (A) Distributed edge networks are involved in sustained attention to visual and auditory tasks. Colored cells reflect pairs of regions whose co-activation time course was significantly positively (red) or negatively (blue) related to

the contrast of interest. Note that only significant edges are visualized. (B) Overlapping edges were involved in sustained attention during both visual-relevant and auditory-relevant runs. Numbers within cells represent the number of edges determined to significantly relate to the sustained attention measure in both visual-relevant and auditory-relevant tasks. Edge networks are summarized using canonical cortical and subcortical networks (Yeo et al., 2011; Tian et al., 2020). Stars (*) in B indicate networks with greater edge involvement than expected by chance.

*Edge networks are not predicted by univariate activity*

Does an edge-based GLM approach offer new insight into neural mechanisms of sustained attention fluctuations? Or, are the co-fluctuations observed to track sustained attention simply byproducts of activity in regions previously reported to underscore attentional fluctuations? To test these questions, we compared networks obtained from our edge-based GLM analysis to a network of theoretical edges predicted by univariate activity, by calculating the signed square root of univariate T-statistics between all pairs of ROIs. This calculation provides a theoretical edge network based on the assumption that, if activity in two ROIs varies as a function of sustained attention, their co-fluctuation should also track sustained attention.

Here, we compare theoretical modality-general edge networks—i.e., edges predicted to be related to sustained attention in both visual and auditory sessions from univariate activity—to observed networks. Theoretical edge maps based on univariate activity predicted large networks of co-fluctuation involved in sustained attention. Despite this, these predicted networks were largely independent of edge networks observed to underlie sustained attentional fluctuations (i.e., the networks in **Figure 12B**). Univariate activity-predicted networks included 4,454 edges stronger during correct omissions and 1,030 edges stronger during lapses across modalities. However, none

of these edges were shared with the networks identified by edge-based GLM analyses (160 edges involved in correct omissions, 271 edges involved in lapses), indicating that theoretical edge predictions missed 100% of observed edges. When contrasting the variance time course against baseline, univariate activity predicted networks of 5,877 edges positively and 1,978 edges negatively related to the variance time course. Of these, 417 positive edges and 33 negative edges were also identified by edge-based GLMs, suggesting that theoretical edge predictions missed 47.2% and 91.8% of observed edges, respectively. In combination, results suggest that edge-based GLMs identify unique neural substrates not predicted by univariate activity alone.

To better understand when edge time series predictions may differ from those made by univariate time series, we simulated pairs of univariate time series and calculated the extent to which their edge time series predicted a theoretical third variable, here representing a behavioral measure of interest. Results are plotted in **Appendix 2 Figure 45**. In general, edges are more likely to predict behavior when univariate time series are both strongly related to one another and strongly related to the behavior of interest. However, as noise in univariate time series increases, these relationships must be strong for the edge time series to be predictive.

*Edges related to sustained attention lapses also track subjective engagement*

Lastly, we tested whether the modality-general networks involved in sustained attention during a controlled task generalized to predict engagement during naturalistic narrative stimuli. Narrative engagement, while related to sustained attention, was measured subjectively and continuously by participants themselves as they watched or listened to movies and a podcast. Successful generalization of edge networks requires that they capture features of attentional fluctuations that are not task-dependent.

Using a leave-one-subject-out approach, we calculated the network strength time course in modality-general edges while participants viewed one silent movie (*Croissant*), one podcast (*Pie*), and two audio-visual movies (*Cake* and *North by Northwest*) in the fMRI scanner. We compared this predicted neural time course to the mean time course of subjective engagement provided by participants using Pearson's correlation. Significance was determined by comparing the average observed prediction (*r*-values) to a null distribution created by circle-shifting the predicted edge time courses within subjects 10,000 times. Mean engagement time courses are visualized in **Appendix 2 Figure 42**.

Edges related to correct omissions vs. lapses during auditory and visual runs of the avCPT also predicted fluctuations in subjective engagement during all narratives (*p*s<.05; **Figure 13A**). As expected, network strength time courses were positively correlated with ratings of subjective engagement across participants, suggesting that co-fluctuation related to better attentional states during the controlled avCPT was also stronger during moments of increased subjective engagement. Correlations between network strength and subject-level engagement time courses were less reliable, but still significant in both audio-visual movies (**Appendix 2 Figure 43a**).

Modality-general edges related to the variance time course predicted subjective engagement during the podcast (*p*<.001) and one audio-visual movie (*Cake*; $p=4.40*10^{-3}$), but not the silent movie or audio-visual *North by Northwest* (**Figure 13B**). The correlation between subjective engagement and VTC edge network strength time courses was predicted to be negative, because edges in the VTC network are inversely related to sustained attention. When predicting subject-level engagement time courses, correlations were less strong but remained significant in the audio-visual *Cake* (**Appendix 2 Figure 43b**). In combination, results show that edges related

to time-resolved measures of sustained attention during a controlled task also capture fluctuations in attentional engagement in a less-controlled, naturalistic paradigm.

Finally, we also tested whether univariate activity in modality-general regions related to sustained attention during the avCPT also predicted subjective engagement during narratives. Results are reported in **Appendix 2 Figure 44**. Activity in regions related to correct omissions vs. lapses tracked engagement in the podcast and both audio-visual movies. The time course of activity predicted from regions related to the variance time course only tracked engagement during the podcast and in audio-visual *North by Northwest*. These results further support the complementary nature of univariate and edge-based predictions.

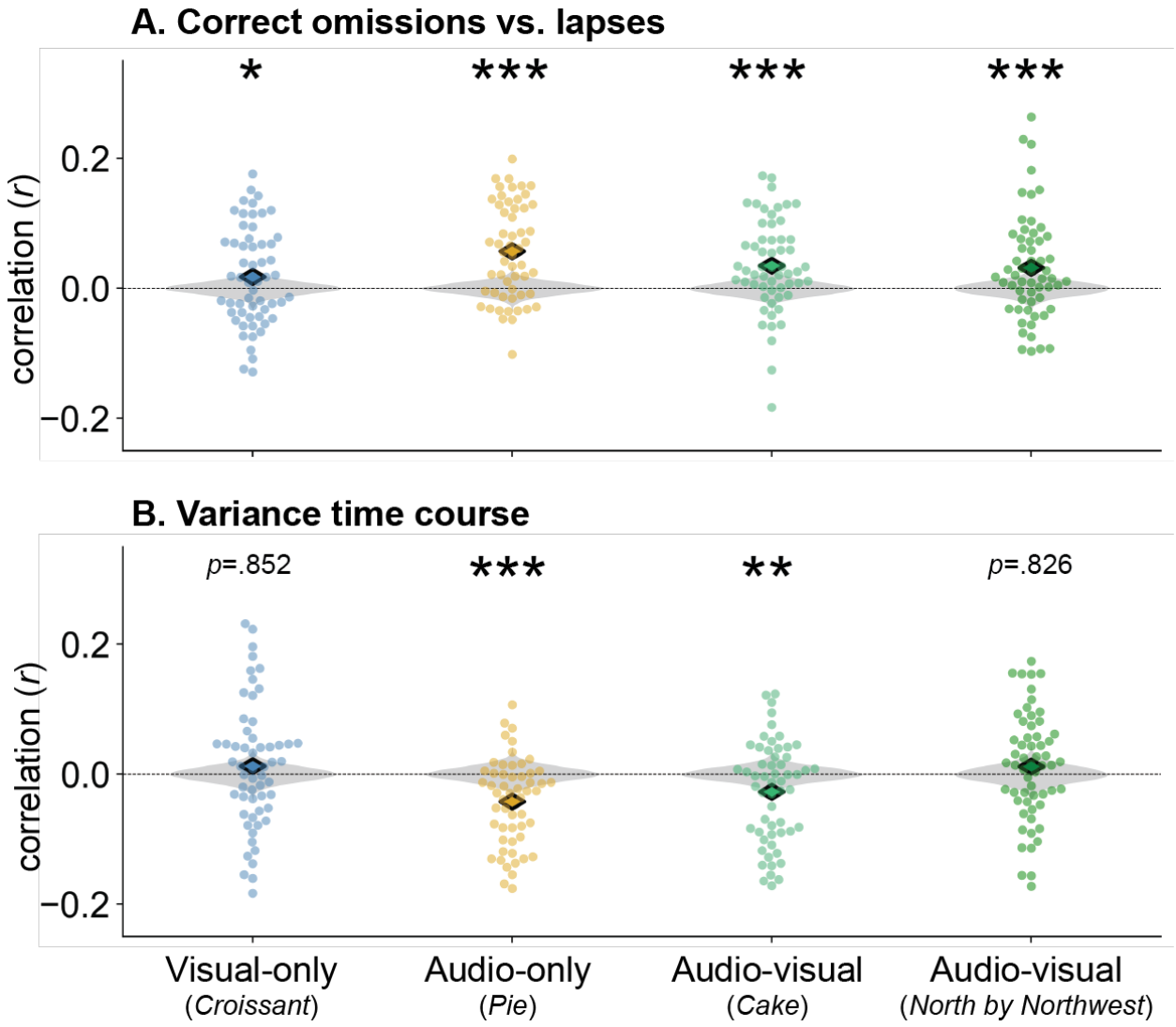



*Figure 13. Sustained attention edge networks track narrative engagement*. Modality general networks related to (A) correct omissions vs. lapses and (B) variance time course fluctuations during the controlled avCPT also captured fluctuations in subjective engagement. Dots reflect individuals' correlation between network strength and mean engagement time courses. Colored diamonds reflect the observed mean correlation across participants. Gray null distributions were created by circle-shifting network strength time courses within participants and recalculating the mean correlation 10,000 times. *p<.05, **p<.01, ***p<.001.

## Discussion

The current study examined the ability of time-resolved edge co-fluctuations to capture dynamic fluctuations of sustained attention from moment-to-moment. Univariate and edge-based GLMs identified a set of core, modality-general regions and networks related to sustained attention regardless of the perceptual modality of the task. Further demonstrating the robustness of edge-based models, modality-general edges also predicted moment-to-moment changes in subjective engagement while participants watched movies and listened to a podcast. Results support the utility of edge time series in the identification of brain networks involved in dynamic attention, cognition, and behavior.

Univariate activity involved in sustained attention largely replicated findings of sensory and attention region involvement from previous work (Fortenbaugh et al., 2018; Jones et al., 2024; Langer & Eickhoff, 2013). Interestingly, the use of an audio-visual CPT in the current study revealed that, during visual-relevant sessions, activity in both visual and auditory regions was increased when participants were performing well, despite auditory stimuli being irrelevant in these sessions. While the interpretation of increased activity is not straightforward, these results are in line with behavioral work finding that processing of irrelevant stimuli (in addition to relevant stimuli) increases under better sustained attentional states during CPTs (Corriveau & Chao et al., 2024; Corriveau et al., 2025b). Additional work is required to examine how processing may be changing at a neural level.

Using edge-based GLMs, we identified modality-general networks whose co-fluctuation tracked trial-to-trial fluctuations in sustained attention. Edge networks were reliable across perceptual modality, mirroring previous work suggesting that networks that predict individual differences in sustained attention contain core modality-agnostic components (Corriveau et al.,

2025a) and demonstrating robustness of edge-based GLM results. Further validating edge-based GLM methodology, networks identified using this time-resolved approach replicate findings from previous methods using functional connectivity averaged over longer time periods. For example, previous work has associated worse sustained attentional states with increased connectivity in the default mode network (Kucyi et al., 2017), in line with our current findings. Default mode network activity has previously been linked with attentional states, though findings differ by task, with greater default activity predicting better sustained attention in controlled tasks (Kucyi et al., 2016; Yamashita et al., 2021) but lower naturalistic engagement (Song et al., 2023).In the current study, edge networks with contributions from the default mode network similarly predicted both attentional lapses and subjective engagement, highlighting the dissociable predictions from univariate and edge-based neural substrates. Additional work using static functional connectivity suggests that increased connectivity within the somatomotor network predicts worse sustained attention across individuals (Corriveau et al., 2025a; Lyu et al., 2025; O'Halloran et al., 2018), a finding we also observed in the current study. This suggests that edge networks, which were defined on temporally-resolved and therefore potentially much noisier patterns of neural activity, capture robust neural substrates underlying sustained attention.

Fewer regions and edges were identified when contrasting correct omissions vs. lapses in auditory-relevant runs relative to visual-relevant runs. While this may reflect key differences between neural activity between perceptual modalities, this may also be the result of worse behavioral performance and possibly greater difficulty during auditory-relevant runs. Errors during auditory-relevant runs may not only reflect lapses in sustained attention but also a failure to correctly categorize a trial. This possibility is also supported by the observation that the edges related to correct omissions vs. lapses were unique from those related to the variance time course

in auditory-relevant runs. Edges in the network tracking correct omissions vs. lapses may reflect not only sustained attention but also processes required by the increased difficulty such as effort, which may not be reflected in the variance time course network. Future work may seek to balance task difficulty across modalities to disentangle these possibilities.

Modality general edge networks also generalized to capture fluctuations in engagement during naturalistic narratives, despite many differences between the controlled avCPT and naturalistic task contexts. Viewing and listening to naturalistic narratives in the scanner was passive, such that participants watched or listened to the stimuli with no experimenter-imposed task goals. While related to sustained attention, narrative engagement may reflect a confluence of phenomena, including participants' internal states and the emotionality of stimuli (Song et al., 2021). Further, narrative engagement was reported subjectively, requiring that participants have accurate insight into their own cognition. Even with these limitations, edge networks trained to predict objective measures of sustained attention, in particular correct vs. incorrect responding, also predicted fluctuations in engagement. Prediction correlation coefficients were small, which may be due to a number of factors, including differences in sampling frequency between measures (predicted edge time series were calculated at every TR whereas engagement was reported much less frequently). Use of non-parametric statistical testing, i.e., circle-shifting predicted edge network time series, ensured that predictions were robust. Previous work using dynamic sliding windows found that a network trained on static functional connections related to sustained attention predicted ongoing engagement during an audio-visual movie but not an auditory-only podcast (Song et al., 2021). The time-resolved approach method used in the current analyses therefore may strengthen the robustness and generalizability of functional networks across task contexts.

Generalization of edge networks to predict engagement during was performed using a leave-one-subject approach, such that brain data from the individual being predicted was not used when calculating the edge network. However, we should note that generalization within a sample collected using the same task and scan paradigms is limited. Future work should extend generalization outside of the training sample to unique tasks, age groups, and scan parameters to further demonstrate the robustness of edge networks.

The current results display the utility of edge-based analyses for the study of neural mechanisms underlying dynamic mental states. Indeed, edge time series are ideal for investigating highly dynamic aspects of attention, cognition, and behavior that may fluctuate at different time scales. Importantly, the edge networks identified using GLMs were not the same as those predicted by univariate activity, despite the fact that edge co-fluctuation time series are mathematically equivalent to the interaction between univariate time courses (Merritt et al., 2024). This suggests that the extent to which the joint fluctuation of two regions tracks sustained attention is unique from those regions' individual relationships with sustained attention. Additionally, in the current study both univariate and edge networks generalized to predict fluctuations in subject engagement during narratives. Recent work, however, has shown that edge co-fluctuations may uniquely capture dynamic cognition when univariate activity does not (Zhang & Rosenberg, 2025). Conversely, it may be that univariate activity more accurately predicts some aspects of cognition and behavior, relative to distributed edge networks. Examining both univariate responses and co-fluctuations using a GLM framework enables the ability to link activity and networks related to phenotypes of interest while still quantifying unique contributions from the two (Jones et al., 2024). Future work seeking to understand what aspects of behavior are better captured by edge co-fluctuations vs. univariate activity may shed light on their unique predictive utility.

We note that our study was able to control for false positives in edge networks by identifying edges that significantly tracked sustained attention in two independent fMRI avCPT runs. Multiple runs provided internal replication, such that we only reported edge networks that were related to sustained attention in both visual-relevant and auditory-relevant runs. Anatomical overlap is one means of identifying functional edges robustly related to common phenotypes across models (Avery et al., 2020; Cheng et al., 2025; Horien et al., 2026; Kucyi et al., 2021; Rosenberg et al., 2018; Wu et al., 2020; Zhang et al., 2025) and has been shown to identify edge networks with superior predictive power (Corriveau et al., 2025b). We should note that in the current study, while models were trained on unique fMRI runs across separate sessions, the participant sample was largely overlapping and may have resulted in greater anatomical overlap than might be observed in unique participant samples. An alternative approach for controlling family-wise error rates may be to use the non-parametric network-based statistic (NBS; Jones et al., 2024; Zalesky et al., 2010). NBS compares the size of a fully-connected network against a distribution of permuted networks generated by dependently permuting rows and columns of a network matrix. This approach may be useful for future work in which internal replication across multiple runs or datasets is not an option.

**Conclusion**

Univariate activity and edge-based GLMs identified brain regions and networks related to moment-to-moment fluctuations in sustained attention. Predictive networks were composed of shared edges regardless of the perceptual modality of the sustained attention task. Further, these modality-general networks were unique from those predicted by univariate activity and captured fluctuations in subjective attentional engagement in a vastly different task context. Together, this

work demonstrates the utility of univariate and edge-based methods for the identification of unique neural substrates underlying dynamic sustained attention.

## Chapter 3 Bibliography


Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., et al. (2020). Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. Journal of Cognitive Neuroscience, 32, 241–255. https://doi.org/10.1162 /jocn_a_01487, PubMed: 31659926

Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. Biological Psychiatry, 57, 1416–1423. https://doi.org/10.1016/j.biopsych.2004.12.005, PubMed: 15950016

Cheng, A., Lichenstein, S., Chaarani, B., Liang, Q., Babaeianjelodar, M., Riley, S. J., et al. (2026). Impulsivity and neuroticism share distinct functional connectivity signatures with alcohol-use risk in youth. Molecular Psychiatry, 31, 953–962. https://doi.org/10.1038/s41380-025-03196-6, PubMed: 40913110

Corriveau, A., Chao, A. F., deBettencourt, M. T., & Rosenberg, M. D. (2025). Recognition memory fluctuates with sustained attention regardless of task relevance. Psychonomic Bulletin & Review, 32, 714–728. https://doi.org/10.3758/s13423-024 -02560-x, PubMed: 39285130

Corriveau, A., James, A. R., deBettencourt, M. T., & Rosenberg, M. D. (2025). Sustained attentional state is a floodlight not a spotlight. Journal of Experimental Psychology: General, 154, 3147–3161. https://doi.org/10.1037/xge0001769, PubMed: 40354291

Corriveau, A., Ke, J., Terashima, H., Kondo, H. M., & Rosenberg, M. D. (2025). Functional brain networks predicting sustained attention are not specific to perceptual modality. Network Neuroscience, 9, 303–325. https://doi.org/10.1162/netn_a _00430, PubMed: 40161982

Corriveau, A., Yoo, K., Kwon, Y. H., Chun, M. M., & Rosenberg, M. D. (2023). Functional connectome stability and optimality are markers of cognitive performance. Cerebral Cortex, 33, 5025–5041. https://doi.org/10.1093/cercor/bhac396, PubMed: 36408606

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, 2712–2723. https://doi.org/10.1093/cercor /bhs261, PubMed: 22941724

Faskowitz, J., Esfahlani, F. Z., Jo, Y., Sporns, O., & Betzel, R. F. (2020). Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. Nature Neuroscience, 23, 1644–1654. https:// doi.org/10.1038/s41593-020-00719-y, PubMed: 33077948

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, PubMed: 29307606

Horien, C., Mandino, F., Corriveau, A., Greene, A. S., O'Connor, D., Shen, X., et al. (2026). Feature consistency in transdiagnostic connectome-based models of sustained attention and autism symptoms. medRxiv, 2026.04.01.26349372. https://doi.org/10.64898/2026.04.01 .26349372, PubMed: 41959777

Jones, H. M., Yoo, K., Chun, M. M., & Rosenberg, M. D. (2024). Edge-based general linear models capture moment-to-moment fluctuations in attention. Journal of Neuroscience,

44, e1543232024. https://doi.org/10.1523 / JNEUROSCI.1543-23.2024, PubMed: 38316565
Kardan, O., Stier, A. J., Cardenas-Iniguez, C., Schertz, K. E., Pruin, J. C., Deng, Y., et al. (2022). Differences in the functional brain architecture of sustained attention and working memory in youth and adults. PLOS Biology, 20, e3001938. https://doi.org/10.1371/journal.pbio.3001938, PubMed: 36542658
Ke, J., Chamberlain, T. A., Song, H., Corriveau, A., Zhang, Z., Martinez, T., et al. (2025). Ongoing thoughts at rest reflect functional brain organization and behavior (p. 2025.08.18.670664). bioRxiv. https://doi.org/10.1101/2025 .08.18.670664
Kucyi, A., Esterman, M., Capella, J., Green, A., Uchida, M., Biederman, J., et al. (2021). Prediction of stimulus-independent and task-unrelated thought from functional brain networks. Nature Communications, 12, 1793. https://doi.org/10.1038/s41467-021-22027-0, PubMed: 33741956
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, U.S.A., 113, 13899–13904. https://doi.org/10.1073/pnas.1611743113, PubMed: 27856733
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, 1831–1840. https://doi.org/10.1093/cercor/bhw029, PubMed: 26874182
Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. Psychological Bulletin, 139, 870–900. https://doi.org/10.1037/a0030694, PubMed: 23163491
Lyu, Y., Corriveau, A., Molla, H., de Wit, H., & Rosenberg, M. D. (2025). Methamphetamine modulates functional connectivity signatures of sustained attention and arousal. bioRxiv. https://doi.org/10.1101/2025.05.20.655181, PubMed: 40475570
Merritt, H., Mejia, A., & Betzel, R. (2024). The dual interpretation of edge time series: Time-varying connectivity versus statistical interaction. bioRxiv. https://doi.org/10.1101/2024 .08.29.609259
Mumford, J. A., Bissett, P. G., Jones, H. M., Shim, S., Rios, J. A. H., & Poldrack, R. A. (2023). The response time paradox in functional magnetic resonance imaging analyses. bioRxiv. https://doi.org/10.1101/2023.02.15.528677
O'Halloran, L., Cao, Z., Ruddy, K., Jollans, L., Albaugh, M. D., Aleni, A., et al. (2018). Neural circuitry underlying sustained attention in healthy adolescents and in ADHD symptomatology. Neuroimage, 169, 395–406. https://doi.org /10.1016/j.neuroimage.2017.12.030, PubMed: 29274748
Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., et al. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. Nature Neuroscience, 19, 165–171. https://doi.org/10.1038 /nn.4179, PubMed: 26595653
Rosenberg, M. D., Hsu, W.-T., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Connectome-based models predict separable components of attention in novel individuals. Journal of Cognitive Neuroscience, 30, 160–173. https://doi .org/10.1162/jocn_a_01197, PubMed: 29040013
Rosenberg, M. D., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., et al. (2020). Functional connectivity predicts changes in attention observed across minutes,

days, and months. Proceedings of the National Academy of Sciences, U.S.A., 117, 3797–3807. https://doi.org/10.1073/pnas .1912226117, PubMed: 32019892
Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., et al. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral Cortex, 28, 3095–3114. https://doi .org/10.1093/cercor/bhx179, PubMed: 28981612
Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66, 44–50. https://doi.org/10.1037/a0025111, PubMed: 21910522
Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. Proceedings of the National Academy of Sciences, U.S.A., 118, e2021905118. https://doi.org/10.1073/pnas.2021905118, PubMed: 34385312
Song, H., & Rosenberg, M. D. (2021). Predicting attention across time and contexts with functional brain connectivity. Current Opinion in Behavioral Sciences, 40, 33–44. https:// doi.org/10.1016/j.cobeha.2020.12.007
Terashima, H., Kihara, K., Kawahara, J. I., & Kondo, H. M. (2021). Common principles underlie the fluctuation of auditory and visual sustained attention. Quarterly Journal of Experimental Psychology, 74, 705–715. https://doi.org/10 .1177/1747021820972255, PubMed: 33103992
Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. Nature Neuroscience, 23, 1421–1432. https://doi.org/10.1038/s41593-020-00711-6, PubMed: 32989295
Unsworth, N., Robison, M. K., & Miller, A. L. (2021). Individual differences in lapses of attention: A latent variable analysis. Journal of Experimental Psychology: General, 150, 1303–1331. https://doi.org/10.1037/xge0000998, PubMed: 40474404
Wu, E. X. W., Liaw, G. J., Goh, R. Z., Chia, T. T. Y., Chee, A. M. J., Obana, T., et al. (2020). Overlapping attentional networks yield divergent behavioral predictions across tasks: Neuromarkers for diffuse and focused attention? Neuroimage, 209, 116535. https://doi.org/10.1016/j .neuroimage.2020.116535, PubMed: 31940476
Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M., Germine, L., Wilmer, J., et al. (2021). Variable rather than extreme slow reaction times distinguish brain states during sustained attention. Scientific Reports, 11, 14883. https://doi.org/10 .1038/s41598-021-94161-0, PubMed: 34290318
Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology, 106, 1125–1165. https://doi.org/10.1152/jn.00338.2011, PubMed: 21653723
Yoo, K., Rosenberg, M. D., Kwon, Y. H., Lin, Q., Avery, E. W., Sheinost, D., et al. (2022). A brain-based general measure of attention. Nature Human Behaviour, 6, 782–795. https://doi .org/10.1038/s41562-022-01301-1, PubMed: 35241793
Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Networkbased statistic: Identifying differences in brain networks. Neuroimage, 53, 1197–1207. https://doi.org/10.1016/j .neuroimage.2010.06.041, PubMed: 20600983

Zamani Esfahlani, F., Jo, Y., Faskowitz, J., Byrge, L., Kennedy, D. P., Sporns, O., et al. (2020). High-amplitude cofluctuations in cortical activity drive functional connectivity. Proceedings of the National Academy of Sciences, U.S.A., 117, 28393–28401. https://doi.org/10.1073/pnas.2005531117, PubMed: 33093200

Zhang, Z., & Rosenberg, M. D. (2025). Brain network dynamics predict moments of surprise across contexts. Nature Human Behaviour, 9, 554–568. https://doi.org/10.1038/s41562-024 -02017-0, PubMed: 39715875

# Chapter 4: Sustained attention's floodlight impacts memory for relevant and irrelevant stimuli

## Recognition memory fluctuates with sustained attention regardless of task relevance

Our sustained attentional state, which fluctuates dynamically over time, influences what we remember. In particular, moments of high attentional engagement during a task are associated with better memory for task information presented in those moments. However, task-related information is often accompanied by information irrelevant to accomplishing the task at hand. What are the consequences of sustained attentional state fluctuations for task-irrelevant information? Specifically, do moments of high sustained attention "sharpen" the focus of selective attention, such that task-relevant stimuli are better processed and remembered while irrelevant stimuli are filtered and forgotten? Or do increases of sustained attention broaden the scope of selective attentional processing to include (or at least reduce filtering of) irrelevant stimuli, leading to better memory? Here, we examine the effects of dynamically changing sustained attentional state on visual memory as a function of task relevance.

Sustained attention dynamics can be probed with continuous performance tasks (CPTs) which require vigilant attention to detect infrequent targets from a set of largely homogeneous stimuli (Mackworth, 1948). In these paradigms, the detection of infrequent stimuli is not perceptually demanding, such that discrimination of frequent and infrequent categories is trivial and errors can be attributed to lapses in attention. Additionally, paradigms that task participants with responding to all frequent-category stimuli and changing a response only to infrequent-category stimuli (not-X CPTs; Robertson et al., 1997; Rosenberg et al., 2013) provide a near-continuous readout of behavior which can be used to approximate sustained attentional state at

every trial. Previous studies investigating fluctuations in sustained attention show that response time (RT) speed and variance predict upcoming lapses in sustained attention during CPTs (Corriveau et al., 2024; deBettencourt et al., 2018; Esterman et al., 2013; Rosenberg et al., 2013; Wakeland-Hart et al., 2022). Specifically, trials in which participants fail to change or inhibit a response to infrequent stimuli, i.e., trials in which participants are likely to be in a decreased sustained attentional state, are preceded by faster and more variable pressing. In this way, RT provides a momentary measure of sustained attention which predicts task performance.

Changes in processing due to fluctuating sustained attentional state impact later memory for task-relevant information (Corriveau et al., 2024; deBettencourt et al., 2018; Madore et al., 2020; Song et al., 2021; Wakeland-Hart et al., 2022). Images presented when individuals are in an engaged attentional state, as indexed by higher CPT accuracy and slower and less variable responses, are better remembered than those shown when people are in a lower attentional state (Corriveau et al., 2024; deBettencourt et al., 2018; Wakeland-Hart et al., 2022). This suggests that processing for task-relevant information may fluctuate with attentional state, affecting how stimuli are remembered later on. Further, salient infrequent stimuli presented amongst a set of homogeneous stimuli may exogenously increase attentional state. Increases in processing for infrequent stimuli may result in better subsequent memory for these stimuli, termed the von Restorff effect (von Restorff, 1933; Wallace, 1965).

Carefully designed studies demonstrate that task-irrelevant stimuli are processed and remembered (Butler & Klein, 2009; Hutmacher & Kuhbandner, 2020; Kuhbandner et al., 2017; Ruz et al., 2005a, b). How does processing of and memory for irrelevant stimuli change as a function of sustained attentional state? Recent work investigated this question using task-relevant and task-irrelevant stimuli presented in auditory and visual perceptual modalities (Corriveau et

al., 2024). During an auditory-visual continuous performance task, participants saw trial-unique images and heard trial-unique sounds simultaneously. They were instructed to make a category frequency judgment on either the images (indoor vs. outdoor scenes) or sounds (natural vs. manmade sounds) and were told that they could ignore stimuli from the irrelevant modality. Results revealed that memory for stimuli from the task-relevant modality positively predicted memory for task-irrelevant items presented at the same time (i.e., in the same attentional state). This suggests that fluctuations in sustained attentional state affect task-relevant and task-irrelevant stimuli similarly. Additional evidence comes from observations that task-irrelevant stimuli presented alongside task-relevant targets are better remembered, termed the "attentional boost effect" (Lin et al., 2010; Swallow & Jiang, 2010). The increase in attention afforded by the presentation of a rare target may increase processing for all stimuli regardless of task relevance, leading to better memory for stimuli.

These findings align with prior work by Esterman et al., (2014) which showed that increases in sustained attentional state resulted in greater neural repetition suppression to irrelevant repeated images, suggesting that sustained attention increased processing for irrelevant information. There is also evidence that other forms of cognitive processing, such as working memory capacity, are increased during heightened sustained attentional states (deBettencourt et al., 2019). These and similar results can also be understood in light of perceptual load theory which posits that processing of task-irrelevant stimuli changes as a function of task load (Lavie, 1995). In conditions of low load, spare attentional capacity can be reallocated to the processing of task-irrelevant information. Do engaged attentional states act similarly, freeing up selective attentional capacity to process task-irrelevant information?

The present study explores the effect of sustained attentional state on selective attention, which has often been characterized in lay terms with a spotlight analogy (Posner, 1980). While much work has debated whether this metaphor appropriately captures all aspects of visual attention (Cave & Bichot, 1999) and instead proposed that multiple mechanisms bias attention in favor of locations, objects, and features based on current goals and stimulus salience (Desimone & Duncan, 1995), the current study will test how sustained attentional fluctuations impact processing in selective attention's scope, which can be diffuse, focused, or divided (LaBerge et al., 1997; Pylyshyn & Storm, 1988; Sperling & Weichselgartner, 1995). Specifically, we test the two opposing hypotheses visualized in **Figure 14**. If sustained attentional state sharpens the focus of selective attention, moments of high attention should see increased processing for stimuli in the focus of selective attention (task-relevant stimuli) and decreased processing for stimuli outside the focus (task-irrelevant stimuli). In this case, we would expect sustained attention to *positively* predict memory for task-relevant stimuli and *negatively* predict memory for task-irrelevant stimuli. Conversely, if sustained attentional state affects processing diffusely, in a manner more akin to a floodlight, moments of high sustained attention should be associated with increased processing of task-relevant stimuli and decreased filtering of task-irrelevant stimuli. If this is the case, we should observe a positive relationship between sustained attention and memory for both task-relevant and task-irrelevant stimuli. While past work found evidence for sustained attention impacting processing in a diffuse manner, this previous work tested memory for irrelevant stimuli presented in separate perceptual modalities (Corriveau et al., 2024). It is not clear whether this pattern would persist when task-relevant and irrelevant stimuli are superimposed images and may therefore compete for limited visual attentional capacity. Clarifying this

relationship would further characterize the multifaceted relationship between sustained attention dynamics and memory.

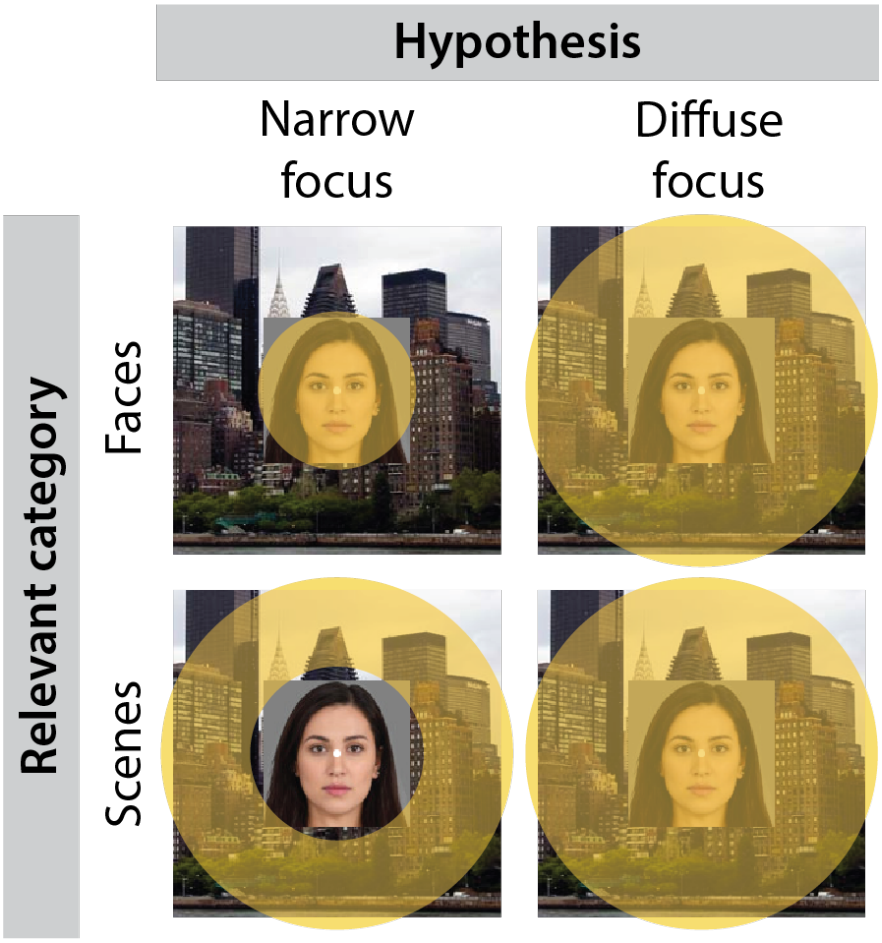



*Figure 14. Predictions for the impact of high attentional state on memory for task-relevant and task-irrelevant stimuli.* A narrow focus hypothesis predicts that individuals in high attentional states will selectively process and remember task-relevant stimuli. Alternatively, the diffuse processing hypothesis predicts that processing during periods of high attention would broaden to include both task-relevant and task-irrelevant information, resulting in higher processing and memory for both

We tested these questions in a set of experiments conducted online. First, we validated the finding that sustained attentional state influences subsequent visual memory in an online experiment. This replication of previous work by deBettencourt et al., (2018) provides an important foundation for further characterizing how sustained attention dynamics influence memory. Next, we probed this relationship by testing how sustained attention affects memory for task-relevant and task-irrelevant images in two independent samples. Results show that, in addition to a positive relationship between sustained attentional state and task-relevant memory, increases in sustained attentional state also positively predicted task-irrelevant item memory. These results suggest that, in this context, sustained attention *decreases* filtering of task-irrelevant information. Thus, stimuli encountered in a higher sustained attentional state are better remembered, regardless of task relevance.

**Experiment 1 methods**

Before testing *how* sustained attention dynamics influence memory, we first sought to replicate foundational work demonstrating that behavioral signatures of sustained attention predict recognition memory in an online sample, for which the testing environment is less constrained and potentially more distracting than it is for in-lab studies. A total of 101 adults ages 18–35 were recruited to participate in an online study hosted by the platform Prolific (www.prolific.com). Participants were excluded for falling 2.5 standard deviations outside the sample mean for a number of criteria, determined a priori. Five participants were excluded based on the number of timed-out attention and memory task trials, two participants for low attention performance ($A'$), one participant for low memory performance ($A'$), and one participant for slow mean reaction time during the memory task. The final sample was composed of 92 participants. A power analysis

using a previously reported effect size of attentional state on memory (preceding RT slope = .18) confirms that the final sample is above the minimum sample size ($N$ = 46) needed to achieve sufficient power (0.8) with a significance level of .05 (Wakeland-Hart et al., 2022).

Participants first performed a continuous performance task (CPT) of sustained attention consisting of a two-alternative forced-choice between frequent-category (90%) and infrequent-category (10%) images. All experiment code was implemented using PsyToolkit (Stoet, 2010, 2017). Stimuli were 550 indoor and 550 outdoor scenes drawn from the SUN image database (Xiao et al., 2010). Participants were presented with a continuous stream of 500 unique images, each lasting 1 second, and were tasked with responding on each trial by pressing the "i" key for an indoor image or the "o" key for an outdoor image.

Following the CPT, participants performed a surprise recognition memory task for scene images. Participants were presented with 200 images from the SUN image database, 100 of which had been presented during the CPT (old) and 100 previously unseen images (new). Of the old images, 50 belonged to the frequent stimulus category and 50 belonged to the infrequent stimulus category. Foil images were category-matched such that 50 belonged to the frequent category and 50 belonged to the infrequent category. Participants rated their recognition memory on a 4-point scale (1 = *definitely new*, 2 = *maybe new*, 3 = *maybe old*, 4 = *definitely old*). Memory for old stimuli was considered correct if a participant reported a judgment of 4 (*definitely old*) and incorrect if a participant provided any other rating, replicating previous work (Corriveau et al., 2024; deBettencourt et al., 2018; Kim et al., 2014; Wagner et al., 1998; Wakeland-Hart et al., 2022). This criterion ensures confident memory judgements and reduces the chance that guesses are counted as correct recognitions (Turk-Browne et al., 2006). Memory trials with no response timed-out after 20 seconds. Timed-out trials are not analyzed.

Finally, to ensure participant compliance, the study concluded with three easy multiple choice questions asking participants to report the kinds of images presented during the task, task instructions for the CPT, and task instructions for the memory task. Participants who failed to answer two of these questions correctly were excluded from the study. No participants were excluded from any of the present studies based on this criterion.

Analysis

Performance for both attention and memory tasks was quantified using the performance measure $A'$. This measure provides a non-parametric quantification of sensitivity using the proportion of hits to false alarms that can be compared against random chance, in this case .5. $A'$ is calculated using the following formula (Grier, 1971):

$$if\ hit > fa, A' = \frac{1}{2} + \frac{(hit - fa) * (1 + hit - fa)}{4 * hit * (1 - fa)}$$

$$if\ fa > hit, A' = \frac{1}{2} - \frac{(fa - hit) * (1 + fa - hit)}{4 * fa * (1 - hit)}$$

As a confirmation that overall memory performance measures were not biased by our accuracy criteria, we also calculated the area under each participants receiver operating characteristic curve (AUC). AUC quantifies an unbiased estimate of memory accuracy by fitting the relationship between false alarm rates and hit rates at all possible criteria (Brady et al., 2023). An equal proportion of false alarm and hit rates is reflected in an AUC value of 0.5 or chance-level performance.

Sustained attention was probed during the CPT using infrequent trials. Errors on infrequent trials reflect lapses in sustained attentional state. RT speed and variance were calculated from a

three-trial window preceding infrequent trials, replicating previous work (deBettencourt et al., 2018, 2019; Decker et al., 2023a, b; Wakeland-Hart et al., 2022; Corriveau et al., 2024; **Figure 15**). RT speed was calculated as the trailing window average of the three correct, frequent trials preceding an infrequent stimulus, after subtracting the overall linear trend (Corriveau et al., 2024; deBettencourt et al., 2018, 2019; Decker et al., 2023a; Wakeland-Hart et al., 2022). This three-trial window was determined a priori. RT variance was calculated using the smoothed variance time course, which quantifies deviation from the mean reaction time over the entire trial (Esterman et al., 2013; Rosenberg et al., 2013). The variance time course is calculated from correct, frequent trials. Missing RTs were interpolated from anchoring trials and the entire time course was smoothed with a gaussian kernel of 8.5 trials. RT variance values were calculated as the mean variance time course values from the three trials preceding an infrequent stimulus (Decker et al., 2023a, b).

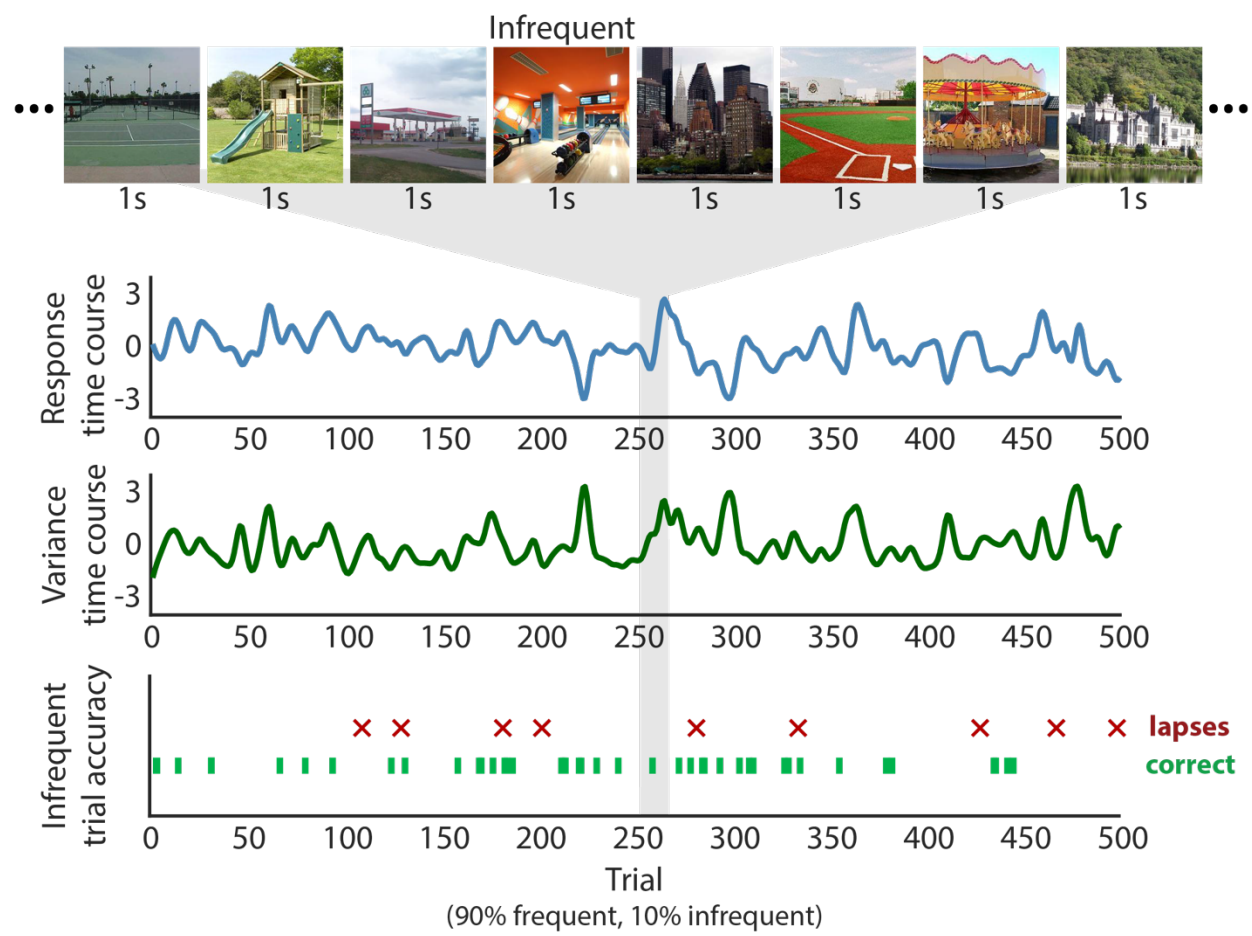



*Figure 15. Measuring sustained attention fluctuations and memory.* **A** In the online CPT, participants viewed streams of frequent (90%) and infrequent (10%) images and made a response indicating whether the image was indoor or outdoor by pressing the 'i' and 'o' keys, respectively. Incorrect responses to infrequent trials reflect lapses in sustained attention. RT and variance time courses quantify speed and variability of pressing throughout the task. Here, the RT time course is smoothed for visualization. **B** Participants completed a recognition memory task following the

CPT, rating memory using a 4-point scale (1 = *definitely new*, 2 = *maybe new*, 3 = *maybe old*, 4 = *definitely old*)

Binomial logistic models were used to test relationships between predictors and performance accuracy, modeled as correct (1) or incorrect (0). Trial-wise CPT accuracy was quantified using categorization judgements on infrequent trials. Trial-wise memory accuracy for previously seen (old) stimuli was considered correct if a stimulus received a memory confidence rating of "definitely old," based on previous work (Corriveau et al., 2024; deBettencourt et al., 2018; Kim et al., 2014; Turk-Browne et al., 2006; Wagner et al., 1998; Wakeland-Hart et al., 2022). Models were built using the *lme4* package in R (Bates et al., 2015). All models included a subject-level random intercept to account for differences in performance between individuals. Additionally, the built-in *bobyqa* optimizer (bound optimization by quadratic approximation; Powell, 2009) was applied to all models. This optimizer did not affect model outputs but ensured convergence for one model which failed to converge when using nonlinear optimization. Data and analysis scripts are publicly available online (https://osf.io/vxuba/).

**Experiment 1 results**

*RT speed and variance uniquely predict lapses in sustained attention*

CPT performance (*A'*) in the final sample was high (*M* = .904, *SD* = .052, 95% CI [.893, .914]) and participants made a category judgement on the vast majority of trials (*M* = .984, *SD* = .017, 95% CI [.980, .987]). We do not report group-level significance above chance because the sample excluded low performers (*n* = 2). However, high performance does ensure the sample is appropriate to further test the relationship between RTs, CPT accuracy, and memory. Therefore,

we next tested whether behavioral variables RT speed and variance predicted lapses in sustained attention during the CPT.

We constructed a logistic model predicting trial-wise CPT accuracy with effects of pretrial RT speed, variance, and their interaction. We only analyzed accuracy to infrequent-category trials because accuracy to frequent-category trials was near ceiling (mean infrequent trial accuracy: 66.7%; mean frequent trial accuracy: 97.6%). RT speed and variance were largely unrelated within-participant (mean $r = .046$, $SD = .250$), mitigating concerns about multicollinearity. We also included a random intercept effect of subject.

*CPT performance ~ RT speed * RT variance + (1|subject)*

Both RT speed ($b = .629$, $SE = .041$, $p < .001$) and variance ($b = -.191$, $SE = .035$, $p < .001$) uniquely predicted lapses in sustained attention such that faster and more variable RTs preceded incorrect presses to infrequent stimuli. The interaction between RT speed and variance also predicted errors ($b = -.064$, $SE = .033$, $p = .049$). These results replicate previous work demonstrating RT speed (deBettencourt et al., 2018; Wakeland-Hart et al., 2022) and variance (Esterman et al., 2013; Rosenberg et al., 2013) predict sustained attentional lapses, as well as a recent observation that they do so uniquely (Corriveau et al., 2024). Therefore, RT measures provide indices of sustained attentional state which may allow us to test its effects on memory.

*Sustained attentional state predicts memory for images*

Recognition memory ($A'$) for stimuli was also high ($M = .712$, $SD = .089$, 95% CI [.693 .730]; **Figure 16A**), demonstrating that participants remembered images presented during the CPT. Memory accuracy as determined by AUC was similarly high ($M = .668$, $SD = .068$, 95% CI

[.654, .682]). Memory performance (*A′*) for infrequent-category stimuli (*M* = .718, *SD* = .132, 95% CI [.691, .745]) was higher than memory for frequent-category stimuli (*M* = .681, *SD* = .092, 95% CI [.662, .700]), $t(91) = -2.81$, $p = 6.14 \times 10^{-3}$), in line with the von Restorff effect (Wallace, 1965). Additionally, CPT performance was positively related to memory (Spearman's *rho* = .218, *p* = .037), such that individuals with higher sustained attention performance had better image memory.

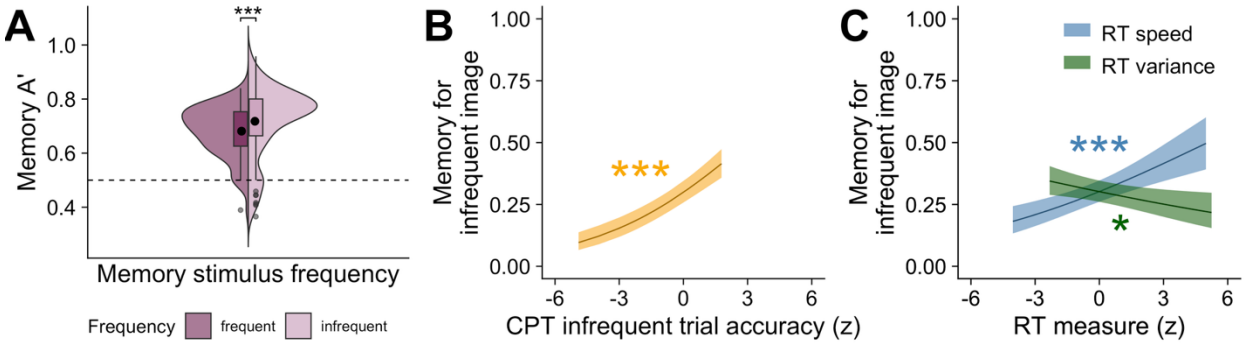



*Figure 16. Experiment 1 results.* **A** Memory for infrequent images was higher than for frequent images. Infrequent stimuli presented during moments of high attentional state as measured by (**B**) CPT accuracy and (**C**) RT speed and variance were better remembered. Models contained a fixed effect of CPT accuracy or fixed effects of RT speed and RT variance, and their interaction, respectively. Both models also included a random subject effect

Finally, we tested whether trial-wise indices of attentional state predicted memory for images. Infrequent-category images require deviation from a prepotent response and therefore provide an assay of sustained attentional state during the CPT. We constructed two logistic models to investigate the effects of sustained attention on memory. The first model contains a fixed effect of infrequent-category CPT trial accuracy and a subject-level random intercept. The outcome

variable is memory accuracy for infrequent images. We did not perform this analysis for frequent-category images because CPT accuracy for these images was near ceiling:

*Memory accuracy ~ CPT accuracy + (1 | subject)*

Lapses in CPT performance significantly predicted lapses in memory ($b$ = .284, $SE$ = .036, $p$ < .001), providing initial evidence that sustained attentional state during encoding impacts which stimuli are later remembered (**Figure 16B**).

Next, we tested whether pretrial RT signatures of sustained attention also predicted memory for infrequent images. A logistic model with fixed effects of RT speed, RT variance, and their interaction was constructed to compare the unique contribution of these effects. A subject-level random intercept was also included in this model:

*Memory accuracy ~ RT speed * RT variance + (1 | subject)*

RT speed ($b$ = .165, $SE$ = .038, $p$ < .001) and RT variance ($b$ = −.084, $SE$ = .035, $p$ = .017) significantly predicted memory for infrequent stimuli, such that stimuli preceded by faster and more-variable pressing were more likely to be forgotten (**Figure 16C**). The interaction between RT speed and variance was not significant ($b$ = −.045, $SE$ = .031, $p$ = .154). These results indicate that behavioral signatures of attentional state predict unique variance in memory for upcoming infrequent images.

**Experiment 2 methods**

Experiment 1 provides an important extension of previous work demonstrating that behavioral signatures of sustained attention predict memory for images. These results provide a foundation to further characterize this relationship in Experiments 2 and 3. In particular, the following experiments test how sustained attention dynamics impact the filtering of—and later memory for—task-irrelevant information. Previous work investigating the consequences of sustained attentional state on memory have focused on the mnemonic fate of task-relevant items. Do fluctuations in sustained attention impact the mnemonic fate of task-relevant and task-irrelevant stimuli similarly? Or do increases in sustained attentional state selectively increase processing for task-relevant stimuli while filtering task-irrelevant stimuli? To probe the effects of trial-wise dynamics of attention on memory, we conducted two related experiments.

In Experiment 2, 201 participants were recruited using Prolific (www.prolific.com). Targeted total sample size was twice that of Experiment 1 ($n$ = 101) because Experiment 2 included two groups with a between-groups manipulation. As in Experiment 1, participants were excluded for performance metrics exceeding 2.5 standard deviations from the mean. Nine participants were excluded based on the number of CPT and memory trials with no response, two for extreme attention performance and two for extreme memory performance. The resulting sample analyzed in Experiment 2 consists of 188 participants.

Participants again performed a 500-trial CPT of sustained attention in which they were tasked with categorizing frequent and infrequent stimuli using a two-alternative forced-choice paradigm (**Figure 17**). However, in this experiment the stimuli comprised face images superimposed on the center of scene images drawn from the SUN image database (Xiao et al., 2010). In Experiment 2, faces were selected from a group of artificially generated images

(https://generated.photos/). Each of these artificial faces was accompanied by a set of metrics describing the generative model's input parameters' association, in [0, 1], with various facial image attributes such as age, sex, emotional expression, and 3D orientation. Faces were selected for our final stimuli on the basis of two criteria: (1) each face must fall within 2.5 standard deviations of the sample mean with respect to each of the above-listed attributes save sex, and (2) each face must fall beyond an absolute threshold of 0.8 with respect to either maleness or femaleness. These criteria were designed to preclude the confounding influence of age, expression, and pose on task performance while supporting unambiguous distinction during the CPT. However, these criteria may have inadvertently disadvantaged memory for our stimuli by restricting image variability (see Experiments 2 and 3, Results). All CPT stimuli included a central fixation dot.

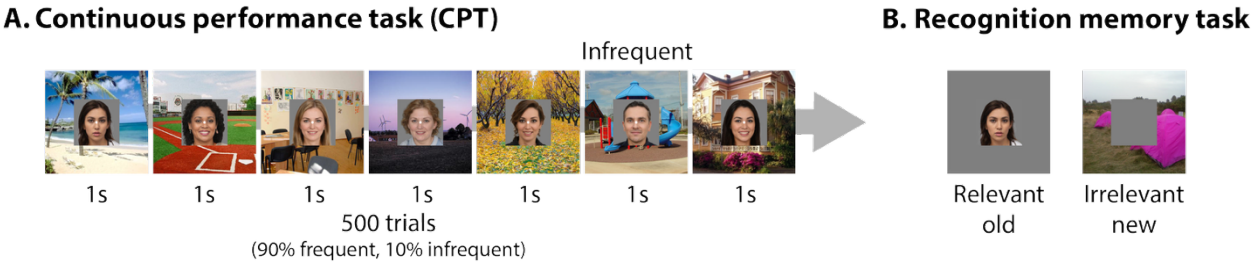



*Figure 17. Experiments 2 and 3 methods.* **A** During the CPT, participants viewed a stream of images composed of faces superimposed on scene images. Participants responded with a button press to either frequent and infrequent faces (depicted) or scenes. **B** Following the CPT, participants performed a memory judgment for relevant and irrelevant stimuli. The recognition memory task contained 400 trials in Experiment 2 and 300 trials in Experiment 3

Task-relevant category (faces or scenes) was manipulated between participants such that participants were assigned to perform a frequent- or infrequent-category judgment on either the face images (male or female) or the scene images (indoor or outdoor). Frequent and infrequent categories were counterbalanced across face and scene images and across participants.

Following the CPT, participants were given a surprise recognition memory task to which they responded with a 4-point scale. Participants were presented with 400 images in total. This was twice as many trials as the memory task in Experiment 1 in order to test memory for both relevant and irrelevant stimulus types. Of these images, 200 had been previously presented during the CPT, 100 as task-relevant (i.e., faces or scenes) and 100 as task-irrelevant stimuli. Within each of these sets, 50 images belonged to the frequent category and 50 to the infrequent stimulus category. The remaining 200 images were category-matched foil images not previously presented.

As in Experiment 1, participants were asked to complete three attention-check questions to ensure compliance after the surprise recognition task. No participants responded incorrectly to more than one attention check question and therefore no participants were excluded on this basis.

**Experiment 3 methods**

Two-hundred participants completed Experiment 3 using Prolific. Six participants were excluded based on the number of attention and memory trials with no response, three for attention performance, and nine for memory performance. The resulting sample in Experiment 3 was 185 participants.

Face images for Experiment 3 stimuli were drawn from a web-scraped and curated database of face photographs, the 10K US Adult Faces Database (Bainbridge et al., 2013). Selection criteria for these face images were less strict than in Experiment 2. Specifically, the age

and head orientation criteria were relaxed to exclude only the youngest-rated faces and face images with averted gaze or head posture. Low-quality images and famous faces were excluded. We did not select images based on facial expression. Finally, images were selected to be maximally unambiguous with regards to sex to facilitate the discrimination task. Differences in selection criteria may have resulted in Experiment 3 face images having more variability in some dimensions (e.g., expression, age, head position) which could result in higher memory.

The memory task for Experiment 3 was designed to probe memory for stimuli that were presented simultaneously, and therefore encoded during the same sustained attentional state. Participants were presented with 300 images, 150 old and 150 foils matched for task relevance and category frequency. This number was lowered from Experiment 2 to reduce overall task length. The previously seen images comprised 75 pairs: 25 task-relevant infrequent stimuli plus their task-irrelevant pairs, 25 task-irrelevant infrequent stimuli plus their task-relevant pairs, and 25 task-relevant frequent stimuli plus their task-irrelevant pairs presented simultaneously during the CPT.

Analysis

Sustained attention and memory performance were again calculated using the sensitivity measure $A'$. AUC values were also calculated for each participant to confirm that memory accuracy criteria did not bias overall performance measures. Pretrial RT speed and variance were calculated using the same methods as Experiment 1. In Experiments 2 and 3, logistic models again included a random intercept for subject and were fit for each experiment separately. Comparing model results across experiments provides a measure of consistency for results across independent datasets.

**Experiment 2 and 3 results**

*CPT performance positively predicts relevant and irrelevant item memory*

CPT performance (*A′*) across participants was high in both Experiment 2 ($M$ = .917, $SD$ = .052, 95% CI [.909, .924]) and Experiment 3 ($M$ = .904, $SD$ = .052, 95% CI [.897, .912]). Additionally, participants responded with a categorization judgement key press on the vast majority of CPT trials ($M_2$ = .986, $SD_2$ = .014, 95% $CI_2$ [.984, .988]; $M_3$ = .983**,** $SD_3$ = .015, 95% $CI_3$ [.981, .986]) suggesting that participants were responding appropriately. Accuracy was .711 ($SD$ = .169) and .674 ($SD$ = .163) for infrequent-category images and .978 ($SD$ = .018) and .974 ($SD$ = .023) for frequent category images in Experiments 2 and 3, respectively.

Memory performance (*A′*) was relatively high for both relevant ($M_2$=.589**,** $SD_2$=.116, 95% $CI_2$ [.572, .606]; $M_3$ = .658**,** $SD_3$ = .078, 95% $CI_3$ [.647, .669]) and irrelevant stimuli ($M_2$ = .551**,** $SD_2$ = .134, 95% $CI_2$ [.532, .570]; $M_3$ = .563**,** $SD_3$ = .125, 95% $CI_3$ [.545, .581]) across experiments. AUC values of memory performance were numerically above chance for both relevant ($M_2$ = .550**,** $SD_2$ = .054, 95% $CI_2$ [.542, .558]; $M_3$ = .597**,** $SD_3$ = .057, 95% $CI_3$ [.589, .606]) and irrelevant stimuli ($M_2$ = .521**,** $SD_2$ = .037, 95% $CI_2$ [.515, .526]; $M_3$ = .517**,** $SD_3$ = .040, 95% $CI_3$ [.511, .522]). Again, we do not statistically compare performance against chance levels because participants were excluded for low performance in these measures. However, overall performance values suggest that participants were not simply guessing and, therefore, that individual trials contain information about sustained attentional state and memory.

Individual differences analyses revealed that overall performance on the CPT was not related to relevant-item memory in Experiment 2 (Spearman's *rho* = .043, $p$ = .561) nor Experiment 3 (Spearman's *rho* = −.033, $p$ = .660). Better overall CPT performance was related to

better memory for irrelevant stimuli in Experiment 2 (Spearman's *rho* = .209, $p = 4.09\times10^{-3}$) but this relationship was not significant in Experiment 3 (Spearman's *rho* = −.021, *p* = .775).

Our main question of interest was how sustained attention fluctuations affected task-relevant and task-irrelevant stimulus memory. Therefore, we next tested the relationship between performance on infrequent CPT trials (i.e., correct presses vs. lapses) and memory for task-relevant stimuli. Both narrow and diffuse models of sustained attention, respectively predicting more and less irrelevant-item filtering during engaged states, would predict that moments of higher sustained attention would lead to greater processing for task-relevant stimuli than moments of low sustained attention. Therefore, we expected a positive relationship, such that correct performance on infrequent CPT trials leads to better memory for images presented during those trials.

We constructed logistic models to test whether sustained attention (i.e., CPT performance) predicted memory. We restricted our analysis to task-relevant, infrequent stimuli, because those trials represent momentary probes of attentional state during the CPT when participants are tasked with changing a prepotent response, whereas accuracy for frequent-category stimuli during the CPT was at ceiling (mean hit rate Experiment 2 = 97.8%; mean hit rate Experiment 3 = 97.4%). The following model was used to test the relationship between sustained attentional state and subsequent memory for individual images:

*Memory for task-relevant, infrequent stimuli ~ CPT accuracy + (1|subject)*

CPT accuracy predicted memory for infrequent, task-relevant stimuli in both Experiment 2 (*b* = .134, *SE* = .029, *p* < .001) and Experiment 3 (*b* = .202, *SE* = .034, *p* < .001; **Figure 18**). This result confirms findings from Experiment 1 and other work (Corriveau et

al., 2024; deBettencourt et al., 2018; Wakeland-Hart et al., 2022) that sustained attentional state predicts memory for task-relevant images. Images presented when sustained attention is high are likely to be remembered, a finding that is predicted by both the narrow and diffuse attentional hypotheses.

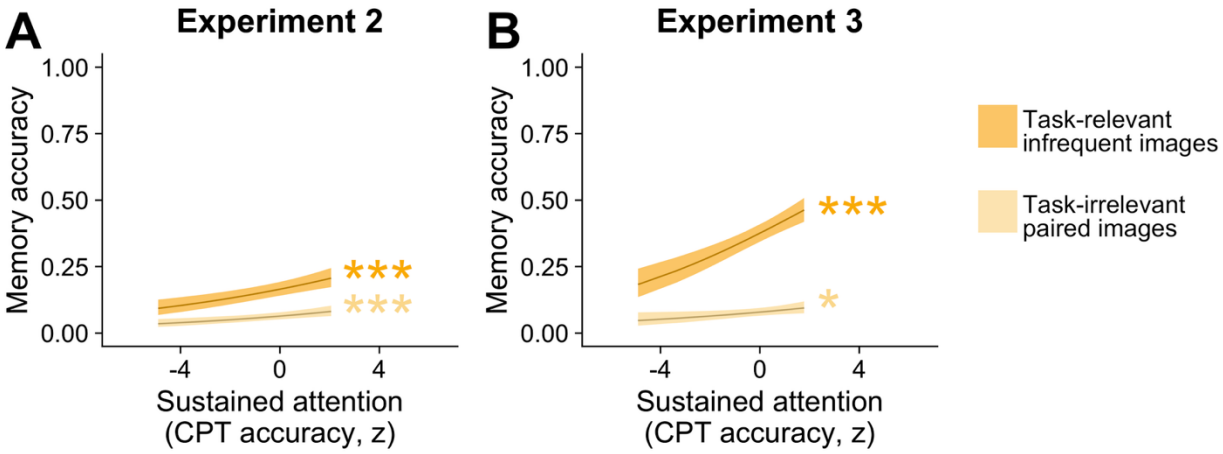



*Figure 18. Sustained attentional lapses predict relevant and irrelevant memory.* Model fits from (**A**) Experiment 2 and (**B**) Experiment 3 predicting memory for task-relevant and task-irrelevant images from CPT accuracy. Models were fit separately for task-relevant infrequent images and their paired task-irrelevant images. Models included fixed effects of CPT accuracy and a subject-level random intercept

The key question is then, what is the mnemonic fate of task-irrelevant stimuli presented in the CPT where selection demands were low? While the narrow and diffuse models of sustained attention predict a positive relationship between sustained attention and memory, they provide different predictions for the mnemonic fate of task-irrelevant stimuli. A narrow model suggests that, during moments of high attention, processing of irrelevant stimuli will be decreased such that they are better filtered. Another possibility is that participants may shift spatial attention to

irrelevant stimuli during an attentional lapse, leading to filtering of task-*relevant* stimuli during lapses. In both cases, we would expect a *negative* relationship between sustained attention performance during the CPT and memory for task-irrelevant stimuli. On the other hand, a diffuse processing model proposes that moments of high attention are characterized by greater processing of all stimuli, such that we would expect a *positive* relationship between CPT performance and memory for irrelevant stimuli.

To distinguish between these narrow and diffuse hypotheses of sustained attention, we next investigated the effect of sustained attention on memory for task-irrelevant stimuli presented alongside infrequent, task-relevant stimuli:

*Memory for task-irrelevant, paired stimulus ~ CPT accuracy + (1|subject)*

We found a positive relationship between CPT accuracy and memory for task-irrelevant stimuli in both Experiment 2 ($b = .127$, $SE = .037$, $p < .001$) and Experiment 3 ($b = .111$, $SE = .051$, $p = .029$; **Figure 18**), suggesting that stimuli presented during moments of high attention were better remembered even if they were not relevant for the task at hand. These results align with the diffuse processing view of sustained attentional state, such that a high sustained attentional state does not increase the filtering of task-irrelevant information, but rather results in more processing for all stimuli regardless of task relevance.

*Stimulus features do not drive the diffuse processing of sustained attention*

One possibility is that these results are driven by the attentional boost effect, such that irrelevant stimuli presented alongside infrequent, relevant stimuli are better remembered (Lin et

al., 2010; Swallow & Jiang, 2010). To test this, we asked whether the frequency of the task-relevant stimuli predicted memory for task-irrelevant stimuli presented at the same time. We find that irrelevant stimuli paired with infrequent, task-relevant stimuli were indeed better remembered in Experiment 3 ($b$ = .318, $SE$ = .062, $p$ < .001) but this was not the case in Experiment 2 ($b$ = −.061, $SE$ = .111, $p$ = .582). Therefore, while the attentional boost effect may contribute to the relationship between sustained attention and task-irrelevant memory in Experiment 3, this effect cannot entirely explain the present results.

Another possibility is that these effects are driven by one condition due to the different categories of stimuli presented. For example, perhaps irrelevant scenes are easily ignored because they surround relevant faces whereas irrelevant faces are always in the center of relevant scenes so may be harder to disregard. We tested whether CPT accuracy predicted relevant-item and irrelevant-item memory within face-relevant and scene-relevant conditions. In face-relevant sessions, CPT accuracy predicted relevant face memory in Experiment 3 ($b$ = .105, $SE$ = .048, $p$ = .030) but not in Experiment 2 ($b$ = $-3.47\times10^{-3}$, $SE$ = .043, $p$ = .936) where face memory was poor (mean $A'$ = .535, $SD$ = .108, 95% CI [.513, .558] for task-relevant faces). In face-relevant sessions, CPT accuracy predicted irrelevant scene memory in Experiment 2 ($b$ = .169, $SE$ = .054, $p$ = $1.62\times10^{-3}$) but not Experiment 3 ($b$ = .126, $SE$ = .068, $p$ = .064). When scenes were the relevant stimulus category, CPT accuracy predicted relevant stimulus memory in Experiment 2 ($b$ = .232, $SE$ = .038, $p$ < .001) and Experiment 3 ($b$ = .287, $SE$ = .047, $p$ < .001). Prediction for irrelevant face memory was positive but non-significant in both Experiment 2 ($b$ = .087, $SE$ = .052, $p$ = .095) and Experiment 3 ($b$ = .092, $SE$ = .080, $p$ = .249). While the strength of predictions was reduced when considering sessions separately, the relationship between CPT accuracy and stimulus memory does not seem to be specific to face-relevant or scene-relevant sessions.

*CPT accuracy predicts confident memory accuracy*

Memory accuracy was determined using confident ratings for old stimuli. In other words, correct memory for an old stimulus constituted a rating of 4 ("definitely old") during the memory task, whereas less confident ratings were marked as incorrect. This decision was based on previous work which utilized the same confidence threshold for memory judgements (Corriveau et al., 2024; deBettencourt et al., 2018; Kim et al., 2014; Wagner et al., 1998; Wakeland-Hart et al., 2022). However, memory judgements were rated using a 1–4 scale which provides insight into memory confidence for each image. Does infrequent CPT trial accuracy predict memory confidence for infrequent stimuli? To test this, we constructed ordinal models predicting memory strength (1–4; 1 = *definitely new*, 2 = *maybe new*, 3 = *maybe old*, 4 = *definitely old*) with a fixed effect of CPT infrequent trial accuracy and a subject-wise random intercept. Models were built using R's *clmm* function from the package *ordinal* (Christensen, 2022). CPT accuracy positively predicted relevant-item memory strength in both Experiment 2 ($b = .091$, $SE = .019$, $p \leq .001$) and Experiment 3 ($b = .157$, $SE = .028$, $p \leq .001$). For irrelevant-item memory, CPT accuracy positively predicted memory strength in Experiment 2 ($b = .046$, $SE = .020$, $p = .020$) but this prediction was not significant in Experiment 3 ($b = 4.22\times10^{-3}$, $SE = .028$, $p = .881$). Therefore, CPT accuracy is related to relevant-item memory confidence but may be less reliable in predicting the memory strength of irrelevant stimuli.

*Memory for a task-relevant image does not predict memory for its task-irrelevant pair*

Since task-relevant and irrelevant stimuli are presented in pairs, each pair was seen in the same attentional state. Therefore, we asked whether memory for task-relevant stimuli predicted memory for their irrelevant pairs. A narrow focus model of sustained attention would predict a

negative relationship, whereby better memory for task-relevant stimuli would be accompanied by better filtering of task-irrelevant stimuli. On the other hand, a diffuse processing model of sustained attention would predict a positive relationship such that better memory for task-relevant stimuli would co-occur with better memory for other information presented simultaneously. To test this, we constructed linear models predicting memory for task-irrelevant stimuli with a fixed effect of task-relevant item memory. This analysis included all pairs of simultaneously presented stimuli tested for memory. Models were constructed for Experiments 2 and 3 separately and intercept varied by subject.

Memory for task-relevant information positively but non-significantly predicted memory for irrelevant paired items in Experiment 3 ($b$ = .047, $SE$ = .030, $p$ = .109) in which the task design maximized the number of task-relevant and task-irrelevant stimulus pairs tested (Experiment 2, $M$ = 55.14, $SD$ = 2.29 pairs tested; Experiment 3, $M$ = 74.76, $SD$ = .80 pairs tested). There was no relationship between memory for task-relevant and task-irrelevant pairs in Experiment 2 ($b$ = .001, $SE$ = .034, $p$ = .965). While the predictions of task-irrelevant item memory from task-relevant memory were positive and therefore in the direction of predictions from the diffuse processing model, this result does not replicate previous findings that task-relevant item memory predicts memory for task-irrelevant information presented simultaneously (Corriveau et al., 2024). On their own, these null findings do not lend strong support to the hypotheses that sustained attention increases or decreases processing of task-irrelevant information.

*Stimulus relevance and frequency predict memory*

What factors, in addition to sustained attentional state, affect memory? For example, prior work on the effect of selective attention would predict that relevant stimuli are better remembered than irrelevant stimuli (Rees et al., 1999). Previous work would also suggest that infrequent stimuli are more salient and therefore may also lead to higher memory performance than observed for frequent stimuli (Corriveau et al., 2024; deBettencourt et al., 2018; Decker et al., 2023b; Wallace, 1965). To test how individual factors influenced memory, we constructed a comprehensive model with sustained attention measures, as well as other stimulus characteristics that may have contributed to memory. For this model, RT speed, RT variance, and their interaction were included as sustained attention measures because they provide a continuous index of attentional state, whereas accuracy was near-ceiling for frequent-category trials.

We first confirmed that, as in Experiment 1, RT speed and variance were reliable predictors of sustained attentional state in Experiments 2 and 3. Within-subject RT speed and variance were again largely unrelated within subjects (Mean $r_2$ = .146, $SD_2$ = .232; Mean $r_3$ = .110, $SD_3$ = .276), suggesting these predictors are not collinear. Linear models predicting CPT performance from RT variables in Experiments 2 and 3 again find that both speed ($b_2$ = .633, $SE_2$ = .031, $p_2$ < .001; $b_3$ = .604, $SE_3$ = .042, $p_3$ < .001) and variance ($b_2$ = −.293, $SE_2$ = .027, $p_2$ < .001; $b_3$ = −.219, $SE_3$ = .036, $p_3$ < .001) predicted lapses in sustained attention. The interaction between RT speed and variance predicted infrequent trial accuracy in Experiment 2 ($b$ = .064, $SE$ = .027, $p$ = .019) but prediction was not significant in Experiment 3 ($b$ = −.046, $SE$ = .035, $p$ = .195). These results validate RT measures as reliable and unique signatures of sustained attention during the CPT, even when task-irrelevant distractors are present.

We next constructed comprehensive mixed effects models to determine what measures, in addition to sustained attentional state, predicted trial-level memory for stimuli. Models included fixed effects of stimulus relevance (irrelevant vs. relevant), frequency (infrequent vs. frequent), and category (scenes vs. faces). RT speed, RT variance, and their interaction were included as measures of sustained attention. CPT accuracy was not included in the model because CPT performance on frequent trials was near ceiling. Models included a random intercept term for individual subjects. Models were fit for Experiments 2 and 3 separately for internal replication of significant predictors. Results are shown in **Table 3**.

*memory accuracy ~ relevance + frequency + category + RT speed * RT variance + (1|subject)*

We observed the expected effect of selective attention such that task-relevant images were better remembered in both experiments. This suggests that participants processed the images relevant for successful CPT performance more than irrelevant images. Stimulus frequency also predicted image memory but the pattern differed between experiments. In Experiment 2, frequent stimuli were remembered better than infrequent stimuli whereas in Experiment 3, we observed the opposite, expected pattern: infrequent stimuli were remembered better than frequent stimuli. Possible explanations of this finding are discussed later. Scenes were remembered better than faces in Experiment 3, but this category effect was not observed in Experiment 2 when controlling for other predictors of memory. Finally, RT variance predicted memory in Experiment 3 such that stimuli preceded by more erratic pressing were more likely to be forgotten. RT speed and the interaction between RT speed and variance did not predict memory in either Experiment 2 or 3. RT measures of sustained attentional state did not significantly predict memory for stimuli in

Experiment 2 when other variables were included in the comprehensive model. While this may be due in part to worse overall memory for face stimuli in Experiment 2, it also suggests that the effects of sustained attention on memory are smaller than other those of predictors (e.g., selective attention) and may be lost when all variables are considered together. These results demonstrate that many factors, including selective and sustained attention, contribute to the mnemonic fate of images.

The finding that infrequent stimuli in Experiment 3 were better remembered aligns with the von Restorff effect and replicates previous work (Corriveau et al., 2024; deBettencourt et al., 2018; Decker et al., 2023b; Wallace, 1965). However, the inverse effect in Experiment 2 across scene and face categories was puzzling. We wondered whether this was a result of the criteria for memory accuracy, such that only confidently-remembered stimuli were considered correctly-remembered. To test this, we refit the models to predict whether stimulus relevance, frequency, and their interaction predicted whether a stimulus received *at least* a memory confidence rating of 3. That is, stimuli were considered correctly remembered if they were reported as "definitely old" or "maybe old," whereas the previous results only considered "definitely old" responses to indicate correct recognition. Here, we will focus on the effects of stimulus frequency on memory. However, full results of this analysis can be found in **Appendix 3 Table 12.** Interestingly, across both experiments, lowering the threshold for correct memory found a reverse in the effect of stimulus frequency on memory, such that frequent stimuli were better remembered in both Experiment 2 ($b = -.343$, $SE = .022$, $p < .001$) and Experiment 3 ($b = -.207$, $SE = .027$, $p < .001$). Therefore, the von Restorff effect may not apply to less-confident judgements of memory, which were more common in Experiment 2, $t(29333) = -14.89$, $p < .001$.

| | **Experiment 2** | | | **Experiment 3** | | |
|---|---|---|---|---|---|---|
| | coef. | SE | sig. | coef. | SE | sig. |
| Relevance *(irrelevant vs relevant)* | -.835 | .031 | <.001 *** | -1.601 | .036 | <.001 *** |
| Frequency *(infrequent vs frequent)* | -.101 | .030 | <.001 *** | .282 | .034 | <.001 *** |
| Category *(scenes vs faces)* | .009 | .031 | .777 | .336 | .036 | <.001 *** |
| RT speed | -.004 | .017 | .825 | .032 | .018 | .077 |
| RT variance | .001 | .016 | .959 | -.034 | .017 | .040 * |
| RT speed: RT variance | .014 | .013 | .292 | .021 | .015 | .145 |

*Table 3. Results from models predicting stimulus memory from fixed effects of stimulus relevance, frequency, category, RT speed, RT variance, and the interaction between RT speed and variance*. Models were fit within-experiment. A subject-level random intercept term was included in each model. *** p<.001, ** p<.01, * p<.05

**Discussion**

This series of experiments investigated the impact of sustained attentional state on recognition memory for images. We first tested whether behavioral predictors of sustained attention also predicted memory in an online sample. In Experiment 1, infrequent-category images were better remembered, in line with previous work showing better memory for unique stimuli (von Restorff, 1933; Wallace, 1965). Images presented during lapses in attention were more likely

to be forgotten, replicating previous work (Corriveau et al., 2024; deBettencourt et al., 2018; Decker et al., 2023b; Wakeland-Hart et al., 2022) in a sample for whom testing conditions were less controlled than in-laboratory studies. RT speed and variance, which predicted lapses in the CPT, also predicted recognition memory for infrequent stimuli. Images preceded by fast and variable presses were more likely to be forgotten, in agreement with previous findings (Corriveau et al., 2024; deBettencourt et al., 2018; Wakeland-Hart et al, 2022). Results from Experiment 1 demonstrate effects of attentional state on memory replicate in less-constrained online samples.

Subsequent experiments tested the impact of sustained attentional state on memory when images were accompanied by irrelevant stimuli. Two independent samples of online participants ($n$ = 373 total) were presented with both task-relevant and task-irrelevant faces and scenes simultaneously. We tested whether sustained attentional state impacted memory for relevant and irrelevant memory in the same direction, as predicted by a diffuse processing view of sustained attention in which engaged attention increases processing of irrelevant items, or opposite directions, as predicted by a narrow focus or tradeoff view of sustained attention in which engaged attention increases filtering of irrelevant items. To do so, we constructed logistic models predicting memory for infrequent task-relevant and task-irrelevant items from sustained attention measured using CPT performance accuracy. Stimuli presented during moments of high attentional state as quantified by accurate CPT performance were better remembered, regardless of task relevance, whereas stimuli presented during moments of low attention were more likely to be forgotten. In other words, attentional state impacted memory similarly for both task-relevant and task-irrelevant stimuli, lending support to the diffuse processing view of sustained attentional state. This view suggests that sustained attention does not enhance selective attentional filtering of task-irrelevant

information when task selection demands are low, but rather supports processing of both task-relevant *and* task-irrelevant stimuli.

Previous work shows that recognition memory is impacted by whether information is attended (Rees et al., 1999; Ruz et al., 2005a) and how attention is directed (Uncapher & Wagner, 2009). The current findings expand on this work by demonstrating that effects of selective attention are impacted by sustained attention dynamics: when sustained attention is engaged, task-irrelevant information is filtered less and remembered better. Related work has investigated differences in recall for task-irrelevant stimuli. The attentional boost effect describes a phenomenon in which irrelevant stimuli paired with a rare task-relevant target are better remembered (Lin et al., 2010; Swallow & Jiang, 2010). Salient targets may be accompanied by an increase in sustained attentional state which may result in the enhanced processing and memory observed. Here, we found evidence for the attentional boost effect in Experiment 3 but not in Experiment 2, suggesting the relationship between sustained attentional state and irrelevant-item memory is not fully driven by stimuli paired with salient targets. Rather, sustained attentional state may fluctuate throughout the task due to a number of reasons in addition to salient stimuli, leading to changes in stimulus processing. Perceptual load theory (Lavie, 1995) provides another framework for understanding the current results. Specifically, previous work has shown that processing of irrelevant items varies as a function of how perceptually demanding task goals are, such that lower task-relevant demand results in increased processing for irrelevant information. Here, high sustained attentional state may be analogous to low perceptual demand, allowing spare capacity for processing and encoding of task-irrelevant stimuli.

Our findings suggest that sustained attention does not affect processing specifically for relevant items in the focus of selective attention, as might be suggested by previous work

characterizing selective attention with a spotlight analogy (Posner, 1980) or by other hypotheses such as biased competition theory positing that bottom-up salience and top-down goals drive attention to behaviorally relevant objects, features, and locations (Beck & Kastner, 2014; Desimone & Duncan, 1995; Vecera & Behrmann, 2001). In the CPT, neither bottom-up salience nor top-down goals should prioritize irrelevant-item processing during moments of engaged relative to disengaged sustained attention. Rather, sustained attentional state may modulate the amount of overall processing that can occur at a given time, including processing of task-irrelevant information. Previous work has similarly questioned the aptness of a spotlight analogy for visuospatial attention (Cave & Bichot, 1999). In particular, work testing the scope of a visuospatial spotlight has found evidence that the distribution of attention over space as well as the strength of processing within this scope changes over time (Eriksen & St. James, 1986; Jonides, 1983; LaBerge & Brown, 1989; LaBerge et al., 1997; Sperling & Weichselgartner, 1995). The current work tests how the amount of processing, both in selective attention's focus (task-relevant stimuli) and outside its focus (task-irrelevant stimuli), varies as a function of sustained attentional state.

It is important to note that the current study cannot distinguish whether a high sustained attentional state enhances memory and/or a low sustained attentional state impedes memory because we have no measure of participants' baseline memory performance. It is possible that high sustained attentional states allow more overall stimulus processing than moments of low sustained attention and therefore result in enhanced memory representations. Another nonmutually exclusive possibility is that during low sustained attentional states, attentional capacity may be redirected to cognitive processes which impede encoding of stimuli into memory, such as mind wandering (Smallwood & Schooler, 2006), although the link between mind wandering and sustained attentional state is complex (Kucyi et al., 2016, 2017). Given that the current study investigates

relative sustained attention and memory performance within individuals, we can only conclude that the relative differences in processing between high and low sustained attentional states lead to relative differences in memory performance.

The current study also does not speak to whether effects of sustained attention dynamics on selective attention vary with cognitive and/or perceptual demands of the central task. The CPT was designed with low selection demands, such that task-relevant image categorization could be performed without necessarily filtering task-irrelevant stimuli. It is possible, therefore, that results would differ under conditions in which task-irrelevant stimuli interfered with task-relevant goals. Additionally, task difficulty in the current study (i.e., categorization of faces and scenes) was low. Increasing the perceptual or cognitive demands of the central task may affect how irrelevant stimuli are processed (Lavie, 1995; Lavie et al., 2004), and this may change as a function of sustained attentional state. These are important areas for future work.

While stimuli in the current study were visual and always consisted of a central face superimposed on a scene image, sustained attention need not affect only visuospatial processing. If this were the case, we might expect the current results to be driven by scene-relevant sessions where irrelevant faces occlude relevant scenes and therefore are difficult to filter. However, the prediction of memory from CPT accuracy within both face- and scene-relevant sessions demonstrates that sustained attention affects stimulus processing similarly across conditions and spatial arrangements of stimuli. Further, previous work testing the effects of sustained attention on auditory and visual stimuli found similar evidence of diffuse processing when relevant and irrelevant stimuli were presented in separate perceptual modalities (Corriveau et al., 2024). Future work should aim to test the contexts in which attentional fluctuations impact processing in a diffuse manner.

Unsurprisingly, we saw a robust effect of selective attention: memory for relevant images was better than memory for irrelevant images across experiments and stimulus types, suggesting that stimuli are better encoded if they are necessary for a task. A comprehensive model which included predictors of selective attention (stimulus relevance) and sustained attention (RT measures) found evidence that these explain unique variance in image memory. These results highlight that attention is not one process but contains separable components that may uniquely affect cognition (Amengual et al., 2022; Chun et al., 2011). We also saw evidence of the von Restorff effect in Experiment 3 such that infrequent stimuli were better remembered than their frequent counterparts. However, lowering the criteria for correct memory judgments saw a flip of this effect in both experiments, such that frequent stimuli were better remembered. Therefore, the von Restorff effect may be specific to confident memory judgements.

The current results demonstrate that sustained attention, which fluctuates over time, predicts subsequent memory. Importantly, it predicts memory for task-relevant and task-irrelevant stimuli similarly, such that moments of high attention lead to higher memory, whereas lapses in attention were associated with forgetting for both task-relevant and task-irrelevant stimuli. This provides further evidence that sustained attention affects visuospatial processing like a flickering floodlight, varying the capacity for processing of both relevant and irrelevant items in tandem. These findings demonstrate the importance of further testing the impact of sustained attention on information processing and cognition.

**Sustained attentional state is a floodlight not a spotlight**

While daily tasks like commuting to work require selectively attending goal-relevant information from multiple perceptual modalities, our ability to do so fluctuates over time (Esterman et al., 2013). Previous work using visual stimuli demonstrates that the attentional state in which task-relevant information is encountered affects how it is processed and whether it is later remembered (deBettencourt et al., 2018; Decker et al., 2022; Wakeland-Hart et al., 2022). When accomplishing tasks, however, relevant information is often accompanied by information irrelevant to the task at hand. What is the fate of task-irrelevant information that is presented during moments of engaged sustained attention? To test this, we use a novel auditory-visual continuous performance task (avCPT) in which participants are presented simultaneous task-relevant and task-irrelevant information in separate auditory and visual perceptual modalities. We first validate the avCPT by testing whether sustained attention to auditory and visual information is related within-individuals and can be indexed by behavioral signatures. Next, we investigate how sustained attentional state during encoding affects memory for task-relevant and task-irrelevant images and sounds. Do moments of high sustained attention sharpen selective attention's spotlight, prioritizing task-relevant information at the expense of task-irrelevant information? Or do high sustained attentional states broaden processing like a floodlight, increasing processing of all information regardless of task-relevance? The present study aims to characterize sustained attentional state and its influence on subsequent memory judgements as a function of perceptual modality and task-relevance.

*Visual and auditory attention share common mechanisms*

We use the avCPT to investigate consequences of attentional fluctuations for recognition memory. Previous studies of auditory and visual sustained attention find that sustained attention ability, measured using continuous performance tasks (CPTs), is positively related within-individuals across visual and auditory perceptual modalities (Seli et al., 2011; Terashima et al., 2020). Thus, the ability to maintain attention to stimuli from separate modalities seems to rely, at least partially, on some shared mechanism or process. Further, interference from auditory processing during a visual task suggests that attending auditory and visual stimuli relies on a common mechanism (Parmentier et al., 2008; although see Mandal et al., 2022). Neuroimaging studies also point to shared, modality-general neural mechanisms of sustained attention. Studies using electroencephalography demonstrate an event-related potential component—the P300 or P3—is responsive to the detection of infrequent targets during auditory and visual oddball tasks (Linden et al., 1999; Katayama & Polich, 1999). The P300 response reflects a positive increase in activity over parietal areas, and its presence in both auditory and visual oddballs suggests that target-detection in these modalities may involve a shared neural response. Functional MRI studies find further evidence that activity in the parietal lobe, particularly in the supramarginal gyrus and inferior parietal lobule as well as in areas of the frontal lobe, increases in response to visual and auditory infrequent targets (Stevens et al., 2000). Common neural mechanisms subserving vigilant attention make it possible to test the effects of sustained attention state when information is presented across perceptual modalities.

*Reaction time measures continuously track visual sustained attention*

Classically, sustained attention was measured using vigilance tasks such as the Mackworth clock task (Mackworth, 1948), which asked participants to report rare, unusually large movements of clock hands occurring irregularly over long periods (0.5-2 hours). More recent CPTs require participants to attend to repetitive information and report deviations with the press of a button (X-CPTs). CPTs are often designed such that the task of discriminating whether a stimulus reflects a deviation from the frequent category is not a perceptually demanding one. Rather, failure to detect a deviation can be attributed to lapses in sustained attention. However, these tasks provide limited sampling of attentional state by requiring responses to rare targets only. This paradigm has been inverted to require responses to frequent stimuli and response inhibitions to rare targets (not-X CPTs; Robertson et al., 1997; Rosenberg et al., 2013). The frequent-response paradigm provides nearly-continuous behavioral insight into attentional state. Work measuring sustained attention to images has found that attention lapses are preceded by moments of faster and more variable reaction times (RTs; deBettencourt et al., 2018; Decker et al., 2022; Esterman et al., 2013; Rosenberg et al., 2013; Yamashita et al., 2021; Wakeland-Hart et al., 2022). Here, we test whether RT measures predict trial-to-trial sustained attention performance to both images and sounds.

*Better memory for images encountered in high attentional state*

How does sustained attentional state at encoding affect later memory? Previous work found that images encountered in moments of high sustained attention (indexed by slower or less-variable RTs) are better recognized (deBettencourt et al., 2018; Decker et al., 2022; Wakeland-Hart et al., 2022). Does sustained attentional state similarly affect memory for sounds? Further, if images and sounds are presented simultaneously, does an engaged attentional state enhance the spotlight

nature of selective attention, processing the task-relevant stimulus while filtering the task-irrelevant one? Or, are moments of high sustained attention more analogous to a floodlight, increasing processing for all information, including information outside of selective attention's spotlight?

Selective attention is indeed well-characterized by a spotlight analogy, such that stimuli in the focus of attention are processed and prioritized over stimuli outside the spotlight or irrelevant stimuli (Norman, 1968; Posner et al., 1980). However, it is less clear how fluctuations in sustained attentional state affect selection in a task where task-irrelevant stimuli are presented in a separate perceptual modality. Do increases in sustained attentional state enhance filtering, resulting in lower memory for irrelevant stimuli? Alternatively, increased attentional capacity afforded by increases in sustained attentional state may lead to enhanced processing and memory of irrelevant stimuli. Evidence for this alternative comes from work by Esterman et al., (2014) which showed increased processing of irrelevant distractor images during successful "in-the-zone" relative to unsuccessful "out-of-the zone" sustained attentional states.

Another related framework comes from perceptual load theory which posits that task-irrelevant stimulus processing during selective attention tasks varies as a function of perceptual task load (Lavie, 1995). Under this theory, spare attentional capacity from conditions of low perceptual load may allow for greater processing of task-irrelevant information resulting in memory for irrelevant stimuli (Lavie et al., 2009). Conditions of high perceptual load, however, resulted in lower processing and poor explicit memory for irrelevant stimuli (Butler & Klein, 2009; Rees et al., 1999). Importantly, careful methodology demonstrates that irrelevant stimuli are indeed processed and remembered (Butler & Klein, 2009; Hutmacher & Kuhbandner, 2020; Kuhbandner et al., 2017; Ruz et al., 2005a; Ruz et al., 2005b). These results suggest that the

mnemonic fate of irrelevant stimuli is related to the amount of processing during encoding. In the current study, task load, i.e., selection demands, remain constant throughout a session. Instead, we ask how memory for irrelevant stimuli changes when that processing varies not as a function of task load itself, but instead as a function of fluctuating sustained attentional state throughout the task.

The present study compares sustained attention and its consequences for recognition memory across auditory and visual perceptual modalities. We present results from a continuous performance task during which participants are presented with sounds and images simultaneously but are instructed to make a response based on one modality, ignoring the other. To test consistency across modalities and participants, we utilized a within-subjects design in which participants performed the task over two sessions, one in which sounds were the relevant category and one in which images were the relevant category, counterbalanced across participants. Following the task, we tested memory for a randomly-selected subset of the task-relevant and task-irrelevant images and sounds. Results demonstrate that sustained attention performance and rate of RT speeding are trait-like within individuals. Additionally, while both pre-trial RT speed and variability predict upcoming lapses in sustained attention for both visual and auditory stimuli, RT variance predicts recognition memory for images only. Finally, we find support for a floodlight view of attentional state, such that stimulus pairs are remembered or forgotten together, regardless of task-relevance.

## Methods

### *Data collection and exclusion criteria*

Data were collected following protocols approved by the University of Chicago institutional review board. Participants were recruited via the online platform Prolific

(www.prolific.co) to participate in a within-subject study that took place in two parts. Sessions were completed on separate days (mean time between sessions=5.91 days, SD=6.05 days), with each session lasting approximately 28.2 minutes (SD=10.4 minutes). Participants who completed session one were invited to return for a second session via Prolific messaging at least 24 hours after the initial session. Participants were compensated $9.75/hour for their time and received a $2 bonus for completing both study sessions. Experimental code was written using jsPsych (de Leeuw, 2015) and implemented via PsiTurk (v3.2.1; Eargle et al., 2020; Gureckis et al., 2016).

227 participants completed at least one online session during which they performed a task designed to test visual or auditory sustained attention (192 completed a visual session, 179 completed an auditory session). Exclusions were made to ensure that online participants were following task instructions and performed the task continuously without leaving their computer. All exclusions were conducted prior to confirmatory analyses. Participants who responded to fewer than half of all continuous performance task trials (successful task performance requires key presses to 90% of trials) or who failed an attention check were not included in analyses (5 participants excluded in the visual session, 3 participants excluded in the auditory session). Additional participants were excluded from analyses if performance on the avCPT (*A'*) was more than 2.5 SD below group mean task performance (6 excluded in the visual condition, 9 excluded in the auditory condition) or if participants were shown repeat stimuli due to experimenter error (4 excluded in the visual condition, 24 excluded in the auditory condition). Participants were asked to report their biological sex at birth from a multiple-choice menu with the following options: Male, Female, Intersex, None of these describe me, Choose not to respond. They also provided their gender from the options: Man, Woman, Non-binary, Other, Choose not to respond. We report biological sex information here. The final sample included 215 participants (ages 18-35, 111

female, 103 male, 1 chose not to respond) in total: 177 participants in the visual condition, 143 participants in the auditory condition, and 105 participants who completed both sessions of the experiment. Of these 105 participants, 49 completed the auditory task and 56 completed the visual task in session one.

Counterbalancing of stimulus presentation by task-relevant modality order and stimulus frequency resulted in 8 possible conditions. We aimed for a sample of 104 participants (13 per condition) who completed both sessions. However, due to participant drop-out between sessions, data collection continued until at least 13 participants completed each condition, resulting in some conditions containing more than 13 participants. A power analysis confirms that the final sample is greater than the minimal sample size required to achieve sufficient power ($1-\beta=0.8$) and significance level ($\alpha=.05$) for a multiple linear model with a fixed effect of attentional state (preceding RT slope=.18; Wakeland-Hart et al., 2022) on memory, i.e., 46 participants.

*Auditory-visual continuous performance task*

To measure sustained attention performance in both auditory and visual domains, participants performed an auditory-visual continuous performance task (avCPT; **Figure 19**). During the task, images and sounds were presented simultaneously for 1000 ms, followed by a 200 ms inter-trial interval. Participants were instructed to respond to either images or sounds. They were told that the stimuli in the task-irrelevant modality would not be important for the task, but that computer volume should remain on and eyes should remain open to receive compensation.

The task-relevant modality was counterbalanced across sessions, such that participants were randomly assigned to either the auditory or visual condition during session 1 and the opposite condition during session 2. To familiarize participants with the task, a 20-trial practice task was

provided at the start of each session, followed by performance feedback. Participants then completed the full 10-minute task (500 trials). After the avCPT, as an attention check, participants were asked to indicate the types of stimuli they were presented throughout the task (e.g., indoor and outdoor scenes) in a multiple-choice format.

Both auditory and visual stimuli contained a frequent category (indoor or outdoor scene images for visual stimuli, natural or manmade sounds for auditory stimuli) that was presented on 90% of trials. Infrequent category stimuli were presented on the remaining 10% of trials. During the auditory session, participants were instructed to respond to sounds and ignore the images. During the visual session, participants were instructed to respond to images and ignore the sounds. Participants were told to press the spacebar for every stimulus in the frequent category of the task-relevant modality and withhold a button press for infrequent category stimuli. Each stimulus presentation contained a black fixation dot which turned gray when the spacebar was pressed to indicate a response had been made. Frequent and infrequent categories were counterbalanced between sessions such that frequent categories during session 1 served as infrequent categories during session 2. Additionally, stimuli were trial-unique across both sessions, such that no image or sound was repeated across either session.

*Stimulus set creation*

Visual stimuli were naturalistic scene images drawn from the SUN397 image database (Xiao et al., 2010). Images belonged to one of two scene categories—indoor or outdoor scenes. To create a rich and diverse image stimulus set, images were drawn from 50 sub-categories for both indoor (e.g., auditorium, kitchen) and outdoor (e.g., gazebo, skatepark) categories. Scene images were cropped to be square. Images were excluded if they included human figures, unusual borders,

obvious photograph enhancement or editing, or category ambiguity (i.e., did not clearly belong to either the indoor or outdoor category). The resulting visual stimulus set included 1,481 indoor images and 1,420 outdoor images.

Auditory stimuli were drawn from online datasets for sounds (Animal.memozee.com; ESC-50 [Piczak, 2015]; Findsounds.com; Freesoundeffects.com; Google Audioset; Mixkit.co; Zapsplat.com) and curated to belong to one of two categories—natural or manmade. The natural category consisted of 21 experimenter-determined subcategories (e.g., dogs barking, water flowing) while the manmade category was made up of 35 subcategories (e.g., musical instruments, car revving). Sound clips were cropped to be 1000 ms in length. Sounds were curated to be unique from other sounds, non-human, and easily distinguishable as natural or manmade. 600 natural and 605 manmade sounds were included in the final auditory stimulus set.

For stimuli to be trial-unique across sessions, a minimum of 593 stimuli were required per modality (auditory and visual) and category (natural/manmade, indoor/outdoor). This number corresponds to the number of stimuli required when a category was the frequent category (450 CPT trials, 50 memory foil images), stimuli for when the category was the infrequent category (50 CPT trials, 25 memory foil images), and an additional 20 stimuli (18 when the category was frequent and 2 when the category was infrequent) for a 20-trial practice CPT at the start of both sessions. Visual stimuli were curated from a large existing database and, therefore, as many images as possible were used from the image categories selected. Auditory stimuli were curated by hand so stimulus collection stopped when enough unique sounds were found. As a result, there is an imbalance in the probability that an individual was presented with any given stimulus between visual and auditory stimulus sets. We do not expect this imbalance to affect current results.

However, stimulus culling could be performed to balance presentation probability for future studies.

*Recognition memory task*

Following the avCPT, participants were tested for recognition memory of the stimuli presented. Recognition for stimuli in the task-relevant modality (i.e., visual or auditory) was tested first, followed by recognition for the task-irrelevant modality. The memory tasks for task-relevant and irrelevant stimuli included 150 trials each (75 old, 75 new). For each task, memory stimuli consisted of 25 old (i.e., previously seen/heard) stimuli from the frequent category, 25 old stimuli from the infrequent category, 25 stimuli that had been paired with the infrequent category of the opposite modality (always from the frequent category of the modality being tested), and 75 foil stimuli.

Participants reported confidence in their recognition memory on a scale from 1-4 (1=definitely new, 2=maybe new, 3=maybe old, 4=definitely old). The memory task was self-paced such that trials ended after a memory judgment was made. However, memory trials timed-out if no response was made within 20 seconds to prevent participants from pausing the task and resuming after an extended period of time. Timed-out trials were not included in analyses. Images remained on the screen until a memory judgment was made. Sounds were presented at the onset of the trial and participants were able to replay sounds as many times as needed before making a memory judgment. On average, participants replayed a small number of relevant (M=2.38; SD=5.01) and irrelevant sounds (M=2.95; SD=6.10) and replay was not related to memory for sounds.

*Questionnaires*

After consenting to participate in each session, participants also completed two surveys, the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) and 10-item Perceived Stress Scale (PSS-10; Cohen & Williamson, 1988). The PANAS asks participants to report their mood in the current moment by rating a series of 10 positive and 10 negative affect items on a five-point scale ranging from "very little or not at all" to "extremely." PANAS results reflect both positive and negative affect state-level scores for each participant. The PSS survey quantifies trait-level stress and asks about participants' thoughts and experiences over the month prior to completing the survey. Finally, to measure potential changes in affect following the task, participants completed a second, identical PANAS survey at the end of each session. These data are not analyzed here.

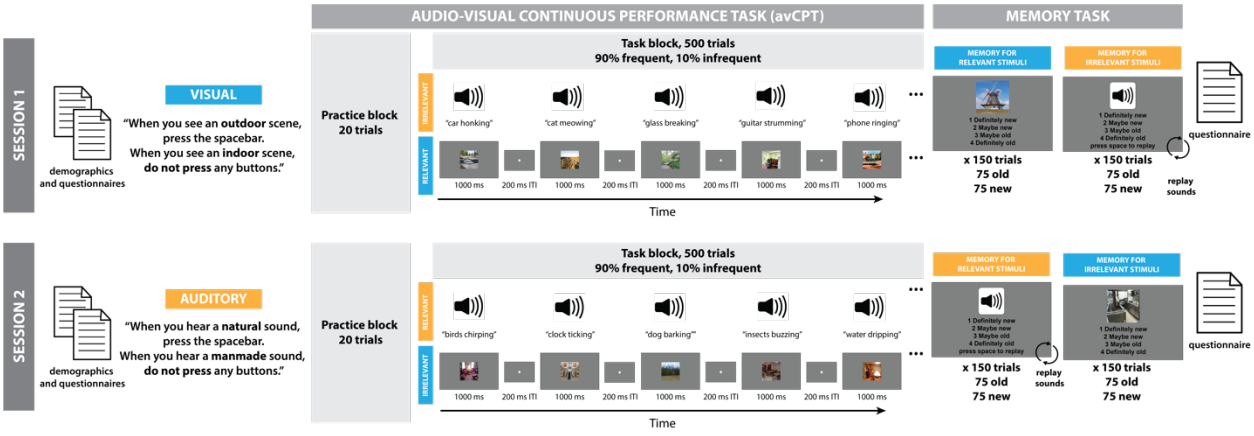



*Figure 19. Experimental procedure.* Participants were invited to complete two online behavioral sessions, each consisting of an audio-visual continuous performance task followed by a recognition memory task. During the avCPT, participants were randomly assigned to respond to auditory stimuli during session 1 and visual stimuli during session 2 or vice versa. Participants were then

tested on recognition memory for both task-relevant and irrelevant stimuli presented during the avCPT using a 4-point scale. See online version for color.

*Analysis*

Overall avCPT performance in each session was quantified using the signal detection theory measure of sensitivity, *A'*, calculated using the following formula (Grier, 1971):

$$if\ hit > fa, A' = \frac{1}{2} + \frac{(hit - fa) * (1 + hit - fa)}{4 * hit * (1 - fa)}$$

$$if\ fa > hit, A' = \frac{1}{2} - \frac{(fa - hit) * (1 + fa - hit)}{4 * fa * (1 - hit)}$$

*A'* provides a non-parametric measure of sensitivity that can be compared against chance-level performance, where chance is 0.5. The change in sustained attention over the course of the avCPT in each session, or the vigilance decrement, was measured as the linear slope of the RT time course across a session. This measure quantifies RT speeding over the CPT, indicating the extent to which pressing became faster and less deliberate.

Higher-frequency changes in sustained attentional state were characterized using two RT measures: RT speed (deBettencourt et al., 2018) and variance (Esterman et al., 2013). RT speed was calculated over three preceding correct frequent-category trials, with faster RTs indicating more automatic, less deliberate, and less attentive responses. If any of the three preceding trials was a commission or omission error, that trial was not included in the calculation of the average pre-trial RT and RT variance. In some cases, this led to the pre-trial RT measures being calculated as the average of two preceding trials or a single trial. However, reanalysis of the data using only

trials with three preceding correct frequent-category trials confirmed that effects are largely unaffected by this analytical decision. Model results from this alternative analysis are included in **Appendix 3 Table 13**.

RT variance was calculated from each session's variance time course (VTC; Rosenberg et al., 2013; Esterman et al., 2013) values of the three preceding correct frequent-category trials. The VTC is calculated as the absolute difference of each RT from the median RT divided by the standard deviation of the RT time course. Only RTs from correct, frequent-category trials were used to calculate the VTC. It quantifies periods of low RT variability, indicating "in-the-zone" attentional states and periods of highly-variable, "out-of-the-zone" attentional states. Both measures were calculated using the detrended RT time course for correct presses to account for linear drift in RT over each run.

Recognition memory for old stimuli was tested using a confidence rating from 1 to 4. In line with previous work, memory was considered correct when participants reported a stimulus was "definitely old" (i.e., a rating of 4; Corriveau et al., 2024; Decker et al., 2022; deBettencourt et al., 2018; Kim et al., 2014; Wagner et al., 1998; Wakeland-Hart et al., 2022) based on work showing that guessing constitutes a large proportion of low-confidence memory reports (Turk-Browne et al., 2006). For new stimuli, a report that the stimulus was "definitely old" was considered incorrect, while any other response was considered correct. Thus, a confident memory judgement is required to make both a hit and a false alarm. Overall performance on the recognition memory tasks was also measured as sensitivity, or *A'*. In rare cases where hit rates and false alarm rates during a memory task were both 0, values were adjusted to the equivalent of a hit and false alarm to a half trial to ensure the resulting *A'* is a real number. In the current context this always

resulted in an *A'* of 0.5, or chance. All main analyses were determined *a priori.* However, alternative analyses are included to confirm that results are robust to analysis decisions.

For all correlation analyses, Spearman's rank correlation values are used to minimize potential effects of outliers in the data. Statistical models were constructed using the *lme4* package in R (Bates et al., 2015). For analyses comparing individual differences between auditory and visual sessions, the sample (*N*=105) of participants who completed both sessions is used. For all other analyses, full samples (visual *N*=177; auditory *N*=143) are used to maximize power.

*Transparency and openness*

Data, experiment code, and analysis code are publicly accessible at https://osf.io/mjy7a/. This study was not preregistered.

**Results**

*CPT performance and attention signatures are consistent across perceptual modalities*

We first tested whether the avCPT is a valid test of sustained attention and can therefore be used to explore the consequences of sustained attention on memory across perceptual modalities. Demonstrating compliance, mean avCPT *A'* was significantly above chance for both auditory (M=.885, SD=.056, 95% CI [.876, .894], t(142)= 82.83, $p<.001$, Cohen's d=6.93) and visual (M=.938, SD=.039, 95% CI [.933, .944], t(176)= 147.99, $p<.001$, Cohen's d=11.12) task sessions (**Figure 20a**). Although these values are inflated based on exclusion criteria requiring participants' avCPT *A'* values fall within 2.5 SD from mean performance, mean avCPT *A*' values were still above-chance prior to removal of low-performing participants (Auditory: M=.869, SD=.085, 95% CI [.857, .882], t(175)=57.20, $p<.001$; Visual: M=.931, SD=.053, 95% CI [.923, .939], t(186)=110.27, $p<.001$). We next tested within-task reliability by correlating *A'* on odd and

even trials within auditory and visual sessions in participants who completed both sessions. Performance was reliable in both auditory (Spearman's *rho*=.738, *p*<.001) and visual (Spearman's *rho*=.683, *p*<.001) sessions. A linear model revealed that avCPT performance was higher during visual sessions (β=.055, SE=$4.31*10^{-3}$, *p*<.001), even after controlling for session number which did not predict avCPT performance in this model (β=$-4.79*10^{-3}$, SE=$4.40*10^{-3}$, *p*=.278). Hit rates on the avCPT were high for both auditory (M=.938, SD=.055, 95% CI [.929, .948]) and visual (M=.990, SD=.022, 95% CI [.987, .994]) sessions and false alarm rates were within the typical range for both auditory (M=.336, SD=.145, 95% CI [.312, .360]) and visual (M=.229, SD=.139, 95% CI [.208, .249]) sessions, based on previous work (Esterman et al., 2013; Rosenberg et al., 2013).

Next, to test whether sustained attention performance was consistent between auditory and visual sessions, we calculated the Spearman's rank correlation of *A'* values between sessions across participants. For all participants who completed both sessions (*n*=105), auditory and visual *A'* was significantly related (Spearman's *rho*=.494, *p*<.001; **Figure 20b**), aligning with prior work suggesting that sustained attention ability is consistent across auditory and visual modalities (Seli et al., 2011; Terashima et al., 2020). Hit rates (Spearman's *rho*=.282, *p*=$3.56*10^{-3}$) and false alarm rates (Spearman's *rho*=.468, *p*<.001) were significantly correlated between sessions, suggesting that pressing tendency is consistent within individuals.

We finally tested whether participants' vigilance decrement during the avCPT was related between auditory and visual sessions. Vigilance decrement was quantified as the slope of the RT time course over each avCPT session. RT time course slopes were positively related (Spearman's *rho*=.232, *p*=.017) such that individuals with a smaller decrement in the visual run were likely to have a smaller decrement in the auditory run. This suggests that individuals' decreases in sustained

attention is trait-like, regardless of sensory modality. However, vigilance decrements were not related to overall *A'* in auditory (Spearman's *rho*=.124, *p*=.141) or visual (Spearman's *rho*=-.038, *p*=.613) runs, suggesting that the rate of RT speeding is not reflective of overall performance.

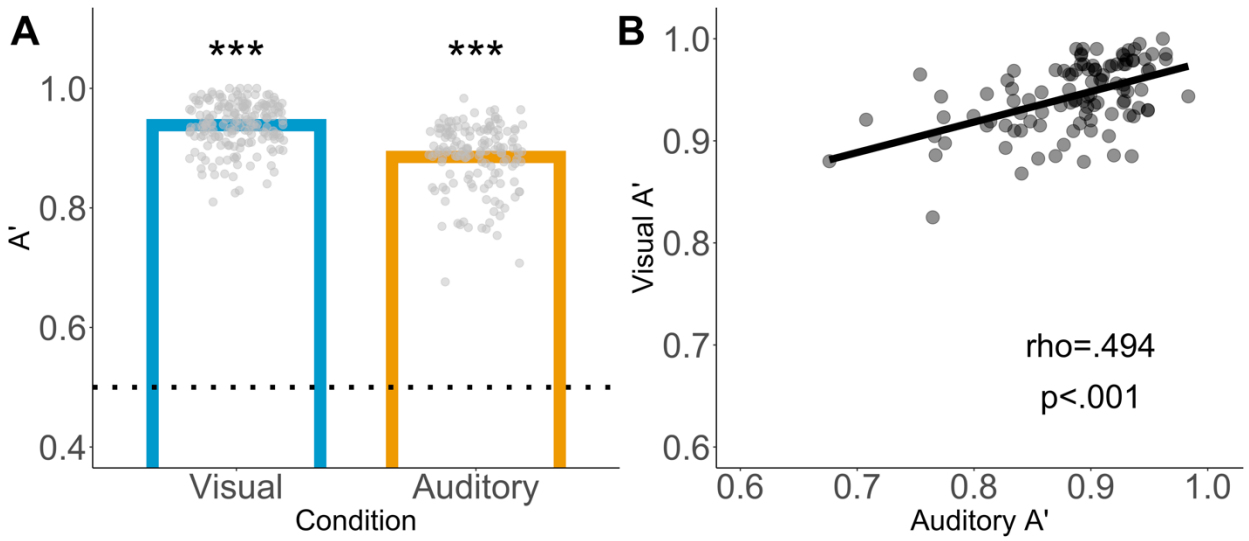



*Figure 20. Sustained attention performance is consistent across modalities*. (A) Average session-level CPT performance across participants was above chance (0.5), ***$p$<.001. (B) Run-wise memory accuracy related across participants for auditory and visual runs. Translucent gray points reflect individual participants' memory performance quantified as sensitivity (*A'*). See online version for color.

*RT measures predict lapses in sustained attention*

While these results suggest that aspects of overall sustained attention performance are stable, we know that attentional state fluctuates within an individual over time and is reliably indexed by transient changes in RT during visual CPTs (Esterman et al., 2013; deBettencourt et al., 2018; Decker et al., 2022; Rosenberg et al., 2013; Wakeland-Hart et al., 2022). Do these

indicators of attentional state predict lapses during both auditory and visual CPTs? We tested whether RT speed and variance during the three trials preceding infrequent stimuli predicted correctly withheld responses. RT speed was calculated from the linearly detrended RT time course using the three trials preceding infrequent, task-relevant stimuli. RT variance was calculated using the unsmoothed VTC for these three preceding trials. On average, within-subject pre-trial RT speed and variance were positively but not perfectly correlated in auditory sessions (mean Spearman's *rho*=.097, SD=.233, 95% CI [.058, .135]) and visual sessions (mean Spearman's *rho*=.247, SD=.231, 95% CI [.213, .281]), suggesting they are not redundant and thus may explain unique variance in sustained attentional state. For both auditory and visual sessions, we constructed a model of performance on infrequent, task-relevant CPT stimuli with fixed effects for RT, RT variance, and their interaction. We included subject-level random intercepts in all models, replicating previous work (Wakeland-Hart et al., 2022). However, results are similar if models include random slope effects. The fixed effects of RT and RT variance are visualized in **Figure 21**.

Replicating previous work in the visual domain (deBettencourt et al., 2018; Wakeland-Hart et al., 2022), RT during the three trials preceding infrequent, task-relevant stimuli predicted whether a response was correctly withheld in both visual (β=.513, SE=.038, $p$<.001) and auditory (β=.147, SE=.036, $p$<.001) sessions, such that lapses were preceded by faster reaction times in both. RT variance also predicted correctly withheld presses in both visual (β=-.238, SE=.030, $p$<.001) and auditory (β=-.172, SE=.027, $p$<.001) runs, with lower variance preceding correctly withheld responses and higher variance preceding lapses. The interaction of pre-trial RT speed and variance was significant in visual sessions (β=-.065, SE=.020, $p$=1.01*$10^{-3}$) but not auditory sessions (β=-.016, SE=.021, $p$=.430). In combination, these results demonstrate that behavioral

signatures of sustained attention previously characterized in visual CPTs explain unique variance in both visual and auditory sustained attention performance.

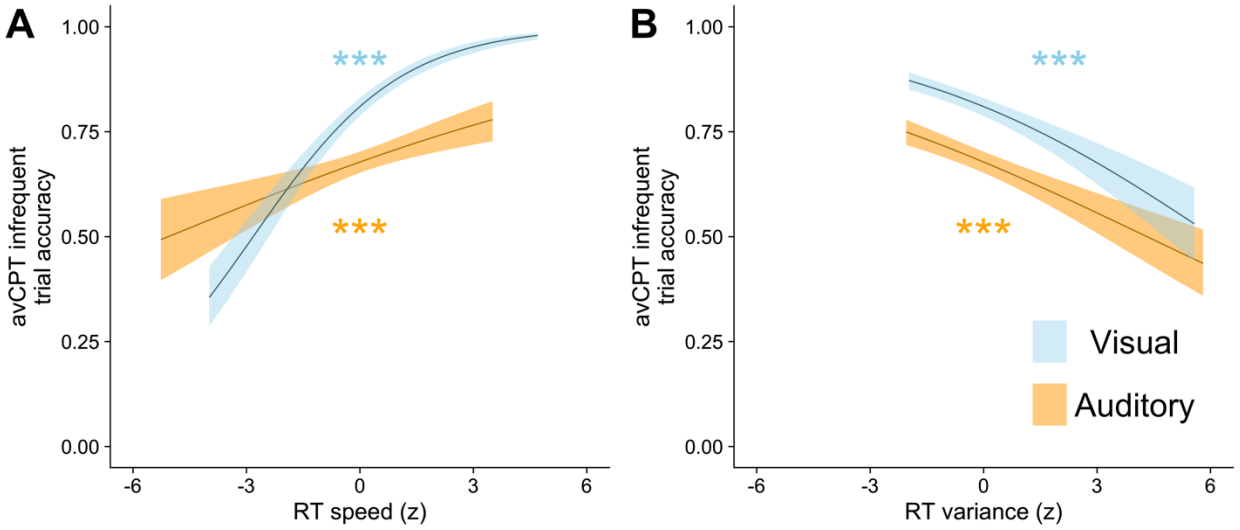



*Figure 21. Response speed and variance predict sustained attentional lapses.* Fixed effects of pre-trial RT speed (A) and variance (B) on CPT performance from a mixed effects model. This model included predictors for pre-trial RT, pre-trial RT variance, and their interaction, as well as a random intercept term for individual subjects. Slower and less variable RTs were related to better performance on infrequent avCPT trials. Shaded areas reflect 95% confidence intervals. ***$p$<.001. See online version for color.

*No evidence of a memory tradeoff for relevant and irrelevant items*

We next evaluated recognition memory performance for task-relevant and task-irrelevant stimuli in both sessions. Mean performance (*A'*) on the memory task was above-chance for images and sounds, regardless of whether stimuli belonged to the task-relevant modality (Auditory M=.603, SD=.095, 95% CI [.587, .618], t(142)=12.87, $p$<.001, Cohen's d=1.08; Visual M=.690,

SD=.089, 95% CI [.677, .703], t(176)=28.41, $p$<.001, Cohen's d=2.14) or the task-irrelevant modality (Auditory M=.555, SD=.098, 95% CI [.541, .570], t(176)=7.48, $p$<.001, Cohen's d=.56; Visual M=.597, SD=.108, 95% CI [.579, .614], t(142)=10.69, $p$<.001, Cohen's d=.89) of a session (**Figure 22a**). Hit rates and false alarm rates for task-relevant and task-irrelevant stimuli are visualized in **Appendix 3 Figure 46**. While main analysis were conducted considering confident memory judgements to old stimuli correct, it is reasonable that less confident memory judgements may reflect true memory. Therefore, we recalculated *A'* such that memory reports to old stimuli of "definitely old" or "maybe old" were considered correct. Memory performance was above-chance using this less stringent threshold (**Appendix 3 Figure 47**), suggesting that results were not largely influenced by this analytical decision.

Recognition memory performance for task-relevant stimuli was marginally related across perceptual modality in participants who completed both sessions (Spearman's *rho*=.192, $p$=.0499; **Figure 22b**). Memory performance for task-relevant and task-irrelevant sounds (Spearman's *rho*=.322, $p$<.001) and task-relevant and task-irrelevant images (Spearman's *rho*=.262, $p=6.83*10^{-3}$) was positively related**.** This suggests that the ability to remember stimuli from a perceptual modality is consistent, regardless of whether the modality is task-relevant or not. To assess the reliability of memory *A'*, we calculated the split-half reliability of memory performance within task-relevant modality by correlating *A'* between even and odd trials for both auditory and visual memory performance. Memory performance was reliable for both visual (Spearman's *rho*=.447, $p$<.001) and auditory sessions (Spearman's *rho*=.280, $p=3.87*10^{-3}$). However, the correlation coefficient of memory performance across modalities approaches the auditory within-task reliability, which provides a theoretical ceiling for possible relationships with other measures.

We also tested whether memory performance for task-relevant and task-irrelevant stimuli was related within a session. A negative relationship between within-session memory performance would suggest a tradeoff between task-relevant and task-irrelevant stimuli, such that one may be prioritized at the expense of the other. However, within-session memory performance was not related when collapsing across perceptual modality (Spearman's *rho*=.063, *p*=.262) and positively related when considering auditory and visual sessions separately (Auditory Spearman's *rho*=.200, *p*=.017; Visual Spearman's *rho*=.167, *p*=.027; **Figure 22c**), providing no evidence of a tradeoff between memory for task-relevant and task-irrelevant stimuli.

Within sessions, avCPT performance was significantly correlated with memory performance for the task-relevant modality in both auditory (Spearman's *rho*=.223, $p=7.40*10^{-3}$) and visual sessions (Spearman's *rho*=.344, *p*<.001), suggesting that better performers in the avCPT also showed higher recognition memory for task-relevant stimuli. Overall avCPT performance during the auditory session was positively correlated with recognition memory for task-irrelevant images (Spearman's *rho*=.165, *p*=.049) but visual avCPT performance was not related to memory for task-irrelevant sounds (Spearman's *rho*=.084, *p*=.264).

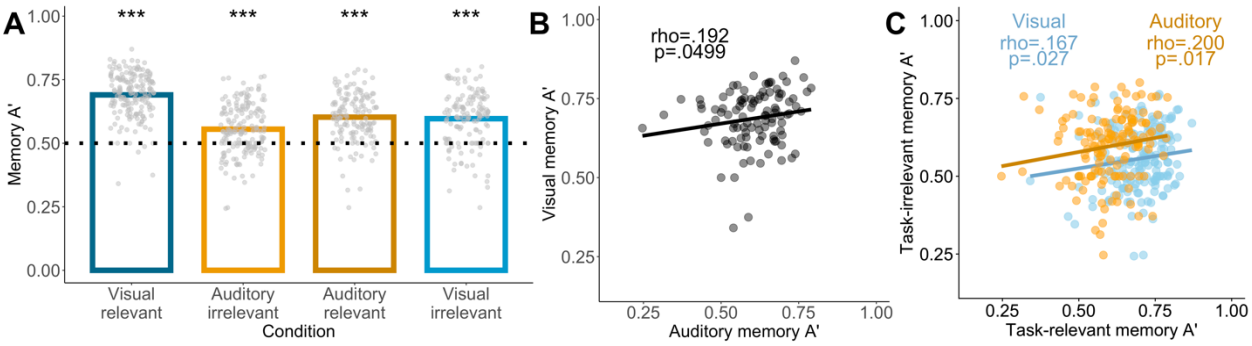



*Figure 22. Memory performance for images and sounds.* (A) Memory performance, quantified as sensitivity (*A'*), was above chance for both relevant and irrelevant auditory and visual stimuli, ***$p<.001$. (B) Within-subject memory performance was marginally related between auditory and visual sessions. (C) Within-session memory for task-relevant and irrelevant stimuli was positively related within-modality. Points represent individual participants' memory performance in a given session. See online version for color.

*Task-relevant and infrequent stimuli are best remembered*

What factors influence whether a stimulus is remembered? For example, is memory performance higher in the second session, when participants might expect a subsequent recognition memory task? Are infrequent-category stimuli more salient and therefore better-remembered? To test these questions, we next investigated memory performance for sounds and images presented during the avCPT. We constructed a comprehensive model that took into account session number, perceptual modality, task-relevance, and stimulus frequency. Further, we included RT predictors of sustained attentional state—pre-trial RT speed and variance—to determine whether attentional state predicted memory above and beyond these other variables. Finally, we included a random effect of subject.

*memory accuracy ~ session + perceptual modality + task-relevance + stimulus frequency + RT speed + RT variance + (1 | subject)*

We observed the expected selective attention effect, such that belonging to the task-relevant modality was the strongest predictor of subsequent memory (**Table 4**). Stimuli from the infrequent category were better remembered than stimuli from the frequent category, showing evidence for the von Restorff effect, which predicts that unique stimuli will be remembered better than homogenous stimuli (Wallace, 1965). Stimuli presented in the first session were better remembered, suggesting that anticipation of a subsequent memory task did not improve memory performance. Images and stimuli preceded by slower RTs were also better recognized.

| Predictor | Coefficient | Standard error | Significance |
|---|---|---|---|
| Session number *1 vs. 2* | .265 | .027 | < .001 *** |
| Perceptual modality *visual vs. auditory* | .083 | .027 | $2.37*10^{-3}$ ** |
| Task relevance *Task-relevant vs. task-irrelevant* | .481 | .022 | < .001 *** |
| Stimulus frequency *infrequent vs. frequent* | .205 | .025 | < .001 *** |
| RT speed | .033 | .011 | $3.52*10^{-3}$ ** |
| RT variance | -.016 | .011 | .150 |

*Table 4. Fixed effects of trial-level recognition memory performance. ***$p$<.001, **$p$<.01.*

*Behavioral correlates of sustained attention predict subsequently forgotten task-relevant images*

We next tested whether measures of sustained attention predicted memory. For these analyses we focus on memory for infrequent, task-relevant stimuli, which have traditionally provided key assays of attention lapses (deBettencourt et al., 2018; Wakeland-Hart et al., 2022). We first tested whether lapses in sustained attention predict recognition memory by constructing logistic models predicting memory from trial accuracy during the avCPT. Models were fit for auditory and visual sessions separately.

$$memory\ accuracy \sim avCPT\ accuracy + (1 \mid subject)$$

Performance during the avCPT significantly predicted recognition memory for both auditory (β=.094, SE=.039, $p$=.017) and visual (β=.168, SE=.036, $p$<.001) stimuli, suggesting sustained attention lapses predict the mnemonic fate of stimuli in both auditory and visual modalities.

Next, we constructed logistic models with predictors of pre-trial RT speed, RT variance, and their interaction to determine whether these indices of sustained attentional state uniquely predicted memory for individual stimuli. Models were constructed for visual and auditory sessions separately and included a random intercept for individual subjects.

$$memory\ accuracy \sim RT\ speed * RT\ variance + (1 \mid subject)$$

RT variance significantly predicted subsequent memory for infrequent, task-relevant images, suggesting that sustained attentional state during encoding affects memory for images (**Table 5; Figure 23**). However, we did not observe a significant relationship between RT speed and memory

for infrequent, task-relevant images nor did RT measures predict memory for sounds. While RT measures predicted lapses in sustained attention, as well as memory in a larger model, RT speed and variability may be noisy measures of sustained attentional state itself and therefore may not capture the effect of attentional state on memory for infrequent, task-relevant images. However, we hypothesized that a more reliable measure of momentary attentional state may be memory itself. To investigate this, we next tested whether memory for a task-relevant stimulus predicted memory for its task-irrelevant pair which was presented in the same avCPT trial.

| Modality | predictor | Coefficient | Standard Error | Significance |
|---|---|---|---|---|
| Auditory | RT speed | .019 | .054 | .718 |
| | RT variance | $4.44*10^{-3}$ | .039 | .909 |
| | RT speed: RT variance | $4.33*10^{-3}$ | .034 | .899 |
| Visual | RT speed | .065 | .048 | .176 |
| | RT variance | -.113 | .038 | $2.66*10^{-3}$ ** |
| | RT speed: RT variance | -.035 | .030 | .239 |

*Table 5. Fixed effects for sustained attention measures on recognition memory for infrequent task-relevant items from joint model.* ***p*<.01.

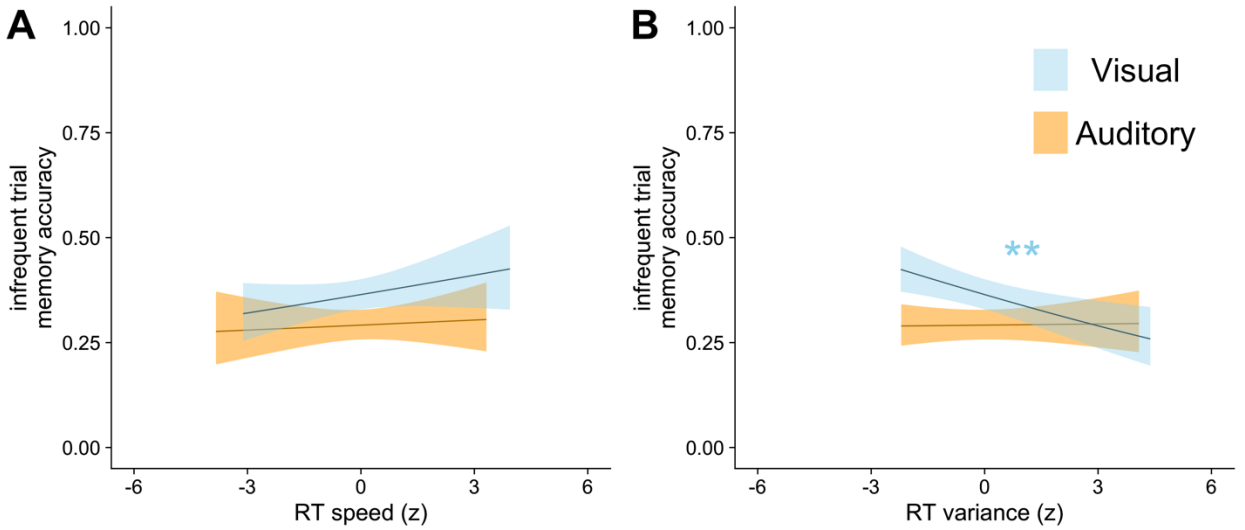



*Figure 23. Response variability predicts image memory.* Fixed effects of RT speed (A) and variance (B) on recognition memory for task-relevant stimuli from a mixed effects model. Model predictors were pre-trial RT, pre-trial RT variance, and their interaction, as well as a random intercept term for individual subjects. Less variable pre-trial RTs were related to better memory for images. Shaded areas reflect 95% confidence intervals. $^{**}p<.01$. See online version for color.

*Memory for task-relevant stimuli predicts memory for paired task-irrelevant stimuli*

Images and sounds in the avCPT were paired randomly and were therefore unrelated. In this context, does sustained attentional state enhance selective attention's spotlight—increasing encoding for relevant stimuli and filtering irrelevant stimuli—or act more like a floodlight—boosting processing for all information, even information outside of selective attention's spotlight? To test this, we asked whether task-irrelevant stimuli presented with a successfully remembered stimulus were more or less likely to be remembered than those paired with forgotten stimuli. A "spotlight" model assuming target-distractor competition would predict that accurate memory for

task-relevant stimulus would *decrease* the likelihood that its task-irrelevant pair would be remembered. A "floodlight" model, on the other hand, would predict that accurate memory for a relevant stimulus (presumably indicating a more engaged attentional state overall) would *increase* the likelihood that its irrelevant pair would be remembered.

We tested whether the proportion of correctly remembered task-irrelevant stimuli differed depending on whether stimuli were presented alongside remembered or forgotten task-relevant stimuli. To maximize power, all stimulus pairs tested for memory—whether stimuli were frequent or infrequent—were used in this analysis. Proportions were calculated by dividing the number of remembered, task-irrelevant stimuli paired with remembered (or forgotten) task-relevant stimuli by the total number of remembered (or forgotten) task-relevant stimuli. Individuals who did not correctly remember any task-relevant stimuli, for whom the denominator of this equation would equal zero, were excluded from this analysis. **Figure 24a-b** visualizes memory performance for task-irrelevant stimuli relative to memory for their task-relevant pairs. Task-irrelevant sounds that were paired with remembered, task-relevant images were more likely to be remembered (memory accuracy M=.281, SD=.209, 95% CI [.250, .313]) than those paired with forgotten images (memory accuracy M=.259, SD=.190, 95% CI [.231, .288]; t(172)=2.34, $p$=.021, Cohen's d=.178). The relationship between the proportion of remembered task-irrelevant images paired with correctly recognized sounds (M=.218, SD=.194, 95% CI [.185, .250]) versus forgotten sounds (M=.202, SD=.191, 95% CI [.170, .234]) was similar but non-significant (t(139)=1.85, $p$=.066, Cohen's d=.156). A mixed effects modeling approach revealed the same pattern: recognition memory for a relevant stimulus positively predicted memory for its paired, irrelevant stimulus (Auditory: β=.054, SE=.026, $p$=.041; Visual: β=.055, SE=.021, $p$=.011; **Figure 24c-d**). This suggests that correctly recognizing a task-relevant stimulus increases the likelihood of recognizing

a paired but irrelevant stimulus presented at the same time, lending support for the floodlight hypothesis of sustained attention.

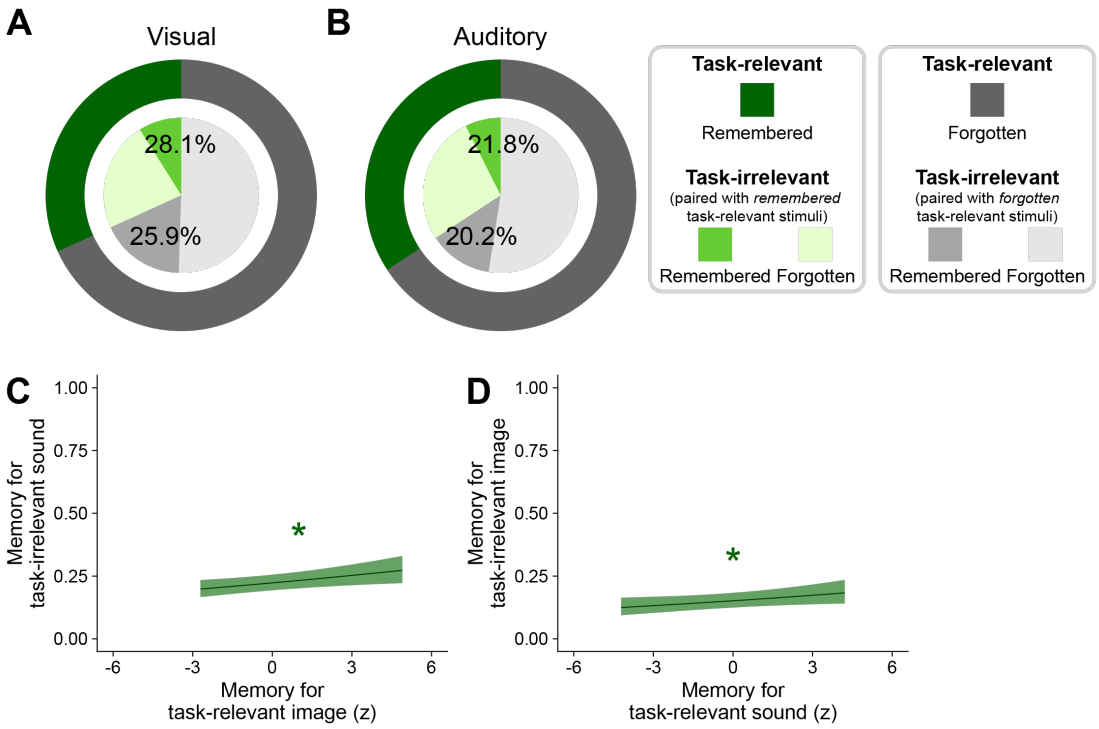



*Figure 24. Memory for relevant and irrelevant pairs.* In both visual (A) and auditory (B) conditions, more task-irrelevant stimuli paired with correctly recognized relevant stimuli were successfully recognized, relative to the proportion of stimuli paired with forgotten relevant stimuli that were successfully recognized. In the inner circle, the darker green wedge constitutes a relatively larger portion than the darker gray wedge, suggesting that remembering task-relevant stimuli was associated with remembering their paired task-irrelevant stimuli, more so than when task-relevant stimuli were forgotten. Percentages represent the percentage of task-irrelevant stimuli paired with remembered (green) or forgotten (gray) task-relevant stimuli. Memory for task-irrelevant sound (C) and task-irrelevant image (D) pairs was predicted by memory for task-relevant

stimuli presented at the same time. Shaded areas reflect 95% confidence intervals. *$p$<.05. See online version for color.

Perhaps this effect is driven by the attentional boost theory, which suggests that, during a target-detection task, the salience of a rare target detected among a stream of frequent stimuli boosts processing and memory for task-irrelevant stimuli presented with the target (Lin et al., 2010; Swallow and Jiang, 2010). If this is the case, we would expect to observe high memory performance for task-irrelevant stimuli that were paired with infrequent, task-relevant stimuli in our data. To test this, we constructed mixed effects models testing whether the frequency of task-relevant items predicted memory for task-irrelevant pairs. We found evidence for the attentional boost theory during visual sessions, such that task-irrelevant sounds paired with task-relevant, infrequent images were better remembered (β=.260, SE=.045, $p$<.001). However, we found no evidence of this effect for task-irrelevant images paired with task-relevant, infrequent sounds (β=$3.36*10^{-4}$, SE=.056, $p$=.995). These results suggest that increased processing during infrequent, task-relevant trials may contribute to the floodlight effect of sustained attention during the visual session. However, the lack of this effect during auditory sessions suggests that the attentional boost theory does not fully explain the memory benefit for task-relevant and task-irrelevant stimuli presented in engaged attentional states.

A significant relationship between memory for task-relevant and task-irrelevant stimuli presented at the same time suggests that the strength of processing varies with sustained attentional fluctuations at encoding. However, other factors may have contributed to this relationship. For example, reactivation of a task-relevant stimulus during the memory task may strengthen memory for other information presented at encoding, as has been demonstrated in previous work (Gardner-

Medwin, 1976; Horner & Burgess, 2013; 2014; Starns & Hicks, 2005). To test whether behavioral signatures of sustained attentional state indeed influenced task-irrelevant memory, we fit separate logistic models testing the relationship between RT speed and task-irrelevant stimulus memory, as well as RT variance as irrelevant-item memory. RT variance did not predict irrelevant-item memory in either auditory (b=-.082, SE=.058, p=.157) or visual sessions (b=.025, SE=.043, p=.569). However, RT speed positively predicted irrelevant-item memory in both auditory (b=.65, SE=.035, p=.064) and visual sessions (b=.081, SE=.028, $p=3.81*10^{-3}$), although this relationship was not significant in the auditory session. This suggests that higher attentional state, measured by longer RTs, was associated with better memory for task-irrelevant stimuli, as predicted by a floodlight view of sustained attentional state.

Further, we tested whether including RT speed as an additional behavioral index of sustained attention state affected the relationship between relevant and irrelevant stimulus memory. A floodlight view of sustained attention could make multiple plausible predictions. One possibility is that irrelevant item memory is positively related to relevant item memory in engaged attentional states when attention is directed to the task. When attention is disengaged, however, irrelevant and relevant item memory may be negatively related (because there are fewer resources to allocate the display overall, inducing competition) or independent of each other (if the breadth and focus of attention is essentially random). This possibility would be reflected in a significant interaction term between attentional state and relevant–item memory such that relevant-item memory should predict irrelevant memory more strongly when stimuli were presented under engaged attentional states. An alternative possibility is that irrelevant and relevant item memory are positively related in both engaged and disengaged states: in engaged states both are processed

simultaneously whereas in disengaged states neither is successfully encoded in memory. We did not have an *a priori* hypothesis about which relationship we would observe.

To address this question we fit logistic models predicting irrelevant-stimulus memory with effects of relevant-stimulus memory, RT speed, and their interaction. Effects of relevant-item memory remained significant in both auditory (b=.058, SE=.0286, p=.029) and visual sessions (b=.054, SE=.022, p=.013). Additionally, RT speed effects were again positive in auditory (b=.064, SE=.035, p=.068) and visual sessions (b=.080, SE=.028, $p=4.21*10^{-3}$), although again this effect did not reach significance in the auditory session. This suggests that irrelevant stimuli presented during moments of high sustained attention are not better filtered. Interaction effects were not significant in either session (auditory b=.037, SE=.034, p=.281; visual b=-.025, SE=.029, p=.383). These results suggest that both sustained attentional and mnemonic effects contribute to successful memory for task-irrelevant stimuli. The lack of an interaction effect would suggest that the relationship between relevant and irrelevant-item memory is not a function of attentional state but, rather, that these effects contribute separately to successful recognition of irrelevant stimuli.

**Discussion**

The current study investigated the consequences of fluctuations in sustained attention on the mnemonic fate of auditory and visual stimuli using the novel avCPT. Results support a perceptual modality-general mechanism of sustained attention indexed by behavioral RT measures. Further, rather than a tradeoff of attention to task-relevant vs. task-irrelevant stimuli, we find that better memory for a relevant image or sound predicts better memory for its simultaneously presented pair. These findings provide insight into the relationship between sustained attention and recognition memory across perceptual modalities.

Despite the potential for distraction, online participants performed auditory and visual sessions of the avCPT successfully. Across modalities, we observed stable individual differences in both overall sustained attention performance and vigilance decrements over time, replicating previous work (Seli et al., 2011; Terashima et al., 2020). RT indices previously shown to predict lapses in visual sustained attention (deBettencourt et al., 2018; Esterman et al., 2013; Rosenberg et al., 2013; Wakeland-Hart et al., 2022) also predict lapses in auditory attention. Furthermore, these measures explain unique variance in lapses, suggesting they may index separable aspects of sustained attentional state. Our results add to accumulating evidence of a modality-general mechanism underlying sustained attention and validate the use of the avCPT for investigating auditory and visual attention.

Memory was above chance for both visual and auditory stimuli, demonstrating that participants successfully encoded images and sounds when they were task-relevant and task-irrelevant. Unsurprisingly, memory was better for task-relevant than task-irrelevant stimuli, providing evidence that selective attention leads to better memory for stimuli in the focus of attention. Memory was also better for images than for sounds, replicating previous work showing poor memory for sounds (Cohen et al., 2009). Infrequent-category stimuli were remembered better than frequent-category stimuli, in line with the von Restorff effect (Wallace, 1965), which predicts better memory for unique items. Finally, recognition memory was better for infrequent images encoded in engaged sustained attentional states, indexed by low RT variance. However, neither RT speed nor variance predicted memory for infrequent sounds, suggesting that RT measures may be noisy signatures of sustained attention.

There was a trending relationship between memory performance across perceptual modalities, such that individuals who remembered more images also remembered more sounds.

Importantly, individual differences analyses also revealed no evidence for a tradeoff between memory for task-relevant and task-irrelevant stimuli within a session. Better memory for task-relevant stimuli did not predict worse memory for task-irrelevant stimuli or vice versa. Thus, contrary to predictions of "spotlight" models of sustained attention, participants who perform well on the memory task for relevant stimuli do not appear to do so at the expense of task-irrelevant stimuli.

To further investigate how moment-to-moment changes in sustained attentional state influence memory, we asked whether memory for a task-relevant stimulus predicted memory for its task-irrelevant pair. In both auditory and visual sessions, memory for the task-relevant stimulus predicted memory for the irrelevant stimulus presented at the same time. When an additional effect of attentional state, RT speed, was included in the model, we observed that both higher sustained attentional state and successful recognition of a paired task-relevant image predicted memory for task-irrelevant sounds. These findings provide additional support for the floodlight view of sustained attentional state, such that moments of higher sustained attentional state enhance processing of task-irrelevant information.

Given the pervasive metaphor of attention as a spotlight, this result may seem counterintuitive. Previous work has argued against a clean spotlight analogy for other forms of attention, such as selective visuospatial attention (Cave & Bichot, 1999). Our results further demonstrate that sustained attention is not well-characterized by a spotlight. They correspond, however, with previous findings of increased—rather than decreased—processing of distractors during in-the-zone sustained attentional states (Esterman et al., 2014). In other words, good performance on a sustained attention task was characterized not by distractor filtering, but by a broad attentional state in which distractors were processed *more*. Results also align with

observations of greater distractor processing during tasks with low perceptual load (Rees et al., 1997; Yi et al., 2004), consistent with the perceptual load hypothesis of distractor processing which predicts increased processing of task-irrelevant information due to spare attentional capacity in low load conditions. Increases in sustained attentional state may be analogous here to low perceptual load conditions, such that additional attentional capacity is available to process distractors. Further, results are congruent with the attentional boost effect, which proposes that salient moments boost processing of both task-relevant and irrelevant information (Lin et al., 2010; Swallow and Jiang, 2010). However, while this effect might suggest that infrequent stimuli drive moments of high attentional state and therefore fully explain the relationship between memory for relevant and irrelevant stimuli, we only saw evidence of this effect during visual sessions and not auditory sessions. Instead, moments of increased processing fluctuate throughout the task, both during salient, infrequent-category trials and repetitive frequent-category trials, shining the "floodlight" of sustained attention on stimuli regardless of task-relevance.

It should be noted that, while task-irrelevant stimuli may have had the potential for distraction in the current study, they did not directly interfere with task goals during the avCPT. The task was designed to test selective attention to images and sounds separately while controlling for potential confounds of presenting images or sounds in isolation. Therefore, selection demand for task-relevant stimuli was low such that accurate performance on the avCPT could be achieved without any need for selective attentional filtering of task-irrelevant information. This may have enabled processing of task-irrelevant stimuli to occur without real detriment to task performance. It is possible that increasing selection demands, i.e., interference between task-relevant and task-irrelevant stimuli, would impact how stimuli are remembered and how memory changes as a

function of sustained attentional state. Future work will test this question by varying the extent to which task-irrelevant stimuli compete with task-relevant goals.

Although the spotlight metaphor does not apply to sustained attention, it may remain appropriate for describing other forms of attention, such as selective spatial attention. The current findings underscore the fact that attention is not a single process (Chun et al., 2011; Amengual et al., 2022) and highlight the importance of disambiguating sustained attention from other attentional components. The floodlight metaphor may also not extend to tasks in which distractor stimuli directly compete with task-relevant goals. That is, in the avCPT, task-irrelevant stimuli were presented in a different perceptual modality and therefore were not in direct competition with task-relevant stimuli for the focus of attention. Further, long trial durations, chosen to allow for clear discrimination of sounds, may have allowed time for encoding of both task-relevant and task-irrelevant information during a single trial. If the task design were such that processing of irrelevant stimuli interfered with task goals, e.g., if relevant and irrelevant stimuli came from the same perceptual modality or if stimulus presentation time was shortened, we might expect a different pattern of results. For example, work by Decker et al., (2023) found evidence for increased distractor processing during attentional lapses when distractors competed with task-relevant stimuli. Further, the avCPT was intentionally designed to have low perceptual load to ensure that errors in category judgements are the result of lapses in sustained attention and not due to difficulty of category discrimination itself. Therefore, it is an open question whether the floodlight metaphor would apply in a more perceptually difficult task of sustained attention. Future work may seek to characterize the extent and limitations of the floodlight metaphor of sustained attention.

Because this study is behavioral in nature, we use behavioral measures of sustained attentional state and memory performance to investigate changes in stimulus processing

throughout the sustained attention task. As a result, it is difficult to disentangle attentional and memory mechanisms in the current results. For example, our results align with previous work demonstrating that paired association memory for a cue and contextual information was dependent, such that memory for a cue-pair association predicts memory for another cue-pair from the same context (Gardner-Medwin, 1976; Horner & Burgess, 2013; 2014; Starns & Hicks, 2005). Indeed, we saw evidence for contributions from both sustained attentional and memory mechanisms in successful memory for task-irrelevant stimuli. Previous work differs from the current study by testing memory using cued retrieval for learned pairs, i.e., participants were explicitly instructed to imagine and memorize associated items. Therefore, all stimuli were task-relevant to some extent. It is also worth noting that previous work tested memory using a cue (Horner & Burgess, 2013; 2014) or tested memory pairs concurrently (Starns & Hicks, 2005) which may have increased the likelihood of observing dependent memory between associated pairs, whereas in the current study, recognition memory was tested using an old-new judgement and tests for relevant and irrelevant stimuli were separated across time (the first half of the memory task contained only relevant-category stimuli, followed by irrelevant-category stimuli in the second half). Therefore, while the effects observed in the current study may be due to similar memory dependence mechanisms, this effect has not been tested with the current methods.

One benefit of the current study design is to provide one explanation for why a relevant stimulus is remembered in the first place, namely the sustained attentional state at encoding. We showed that sustained attentional state predicts recognition memory for task-relevant stimuli, and retrieval of a memory strengthens retention of contextual information presented at encoding (Jonker et al., 2018). Therefore, it may be that the effect of sustained attentional state on task-irrelevant stimulus memory is two-fold. Higher sustained attentional state at encoding may lead to

stronger memory for a task-irrelevant stimulus itself. Further, task-relevant stimuli presented during higher sustained attentional state are more likely to be remembered, reactivating memory for irrelevant stimuli presented at the same time. Future work using neural methods such as fMRI may be better-equipped to inform what mechanisms underlie this relationship.

Our current results challenge the notion that all forms of attention act as a spotlight, enhancing processing for only goal-relevant information. Instead, fluctuations in sustained attention led to similarly increased (or decreased) processing for all stimuli regardless of task-relevance. Future work can interrogate the boundaries of the sustained attentional floodlight within and across perceptual modalities and contexts. Further, consequences of sustained attentional state may extend to other cognitive processes impacted by fluctuations in attention, such as learning. Fully characterizing the extent and effects of sustained attention's floodlight may provide insight into processes involved in sustained attention and its interaction with other forms of attention and cognition.

*Constraints on Generality*

Participants in the current study were adults ages 18-35. While the effects observed may generalize to populations outside this age range, we cannot conclude that results would replicate in all age groups.

**Chapter 4 Bibliography**


Amengual, J. L., Di Bello, F., Ben Hadj Hassen, S., & Ben Hamed, S. (2022). Distractibility and impulsivity neural states are distinct from selective attention and modulate the implementation of spatial attention. Nature Communications, 13(1), Article 4796. https://doi.org/10.1038/s41467- 022-32385-y

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. Journal of Experimental Psychology: General, 142(4), 1323–1334. https://doi.org/10.1037/a0033872

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

Beck, D. M., & Kastner, S. (2014). Neural systems for spatial attention in the human brain: Evidence from neuroimaging in the framework of biased competition. In A. C. Nobre, & S. Kastner (Eds.), Oxford handbook of attention. Oxford University Press. https:// doi.org/10.1093/oxfordhb/9780199675111.013.011

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. Psychonomic Bulletin & Review, 30(2), 421–449. https://doi.org/10.3758/ s13423-022-02179-w

Butler, B. C., & Klein, R. (2009). Inattentional blindness for ignored words: Comparison of explicit and implicit memory tasks. Consciousness and Cognition, 18(3), 811–819. https://doi.org/10. 1016/j.concog.2009.02.009

Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. Psychonomic Bulletin & Review, 6(2), 204–223. https:// doi.org/10.3758/BF03212327

Christensen, R. H. B. (2022). ordinal: Regression models for ordinal data (2022.11-16) [Computer software]. https://cran.r-project.org/ web/packages/ordinal/index.html

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. Annual Review of Psychology, 62(1), 73–101. https://doi.org/10.1146/annurev.psych. 093008.100427

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proceedings of the National Academy of Sciences of the United States of America, 106(14), 6008–6010. https://doi.org/10.1073/pnas.0811884106

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), The social psychology of health: Claremont Symposium on applied social psychology (pp. 31–67). Sage Publications.

Corriveau, A., James, A. R., Jr., deBettencourt, M. T., & Rosenberg, M. D. (2024). Sustained attentional state is a floodlight not a spotlight. PsyArXiv Preprints. https://doi.org/10.31234/osf.io/k9cnm

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior Research Methods, 47(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

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

deBettencourt, M. T., Keene, P. A., Awh, E., & Vogel, E. K. (2019). Real-time triggering reveals concurrent lapses of attention and working memory. Nature Human Behaviour, 3(8), Article 8. https://doi.org/10.1038/s41562-019-0606-6

Decker, A., Dubois, M., Duncan, K., & Finn, A. S. (2023a). Pay attention and you might miss it: Greater learning during attentional lapses. Psychonomic Bulletin & Review, 30(3), 1041–1052. https://doi.org/10.3758/s13423-022-02226-6

Decker, A. L., Duncan, K., & Finn, A. S. (2023b). Fluctuations in sustained attention explain moment-to-moment shifts in children's memory formation. Psychological Science, 34(12), 1377–1389. https://doi.org/10.1177/09567976231206767

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Eargle, D., Gureckis, T., Rich, A. S., McDonnell, J., & Martin, J. B. (2020). psiTurk: An open platform for science on Amazon Mechanical Turk (Version 3.2.1) [Computer software]. Zenodo. https://doi.org/10.5281/ze nodo.3598652

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. Perception & Psychophysics, 40(4), 225–240. https://doi.org/10.3758/ BF03211502

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., 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

Gardner-Medwin, A. R. (1976). The recall of events through the learning of associations between their parts. Proceedings of the Royal Society of London. Series B. Biological Sciences, 194(1116), 375–402. https:// doi.org/10.1098/rspb.1976.0084

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, 75(6), 424–429. https://doi.org/10.1037/h0031246

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An opensource framework for conducting replicable behavioral experiments online. Behavior Research Methods, 48(3), 829–842. https://doi.org/10 .3758/s13428-015-0642-8

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. Journal of Experimental Psychology: General, 142(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. Current Biology, 24(9), 988–992. https://doi.org/10.1016/j .cub.2014.03.012

Hutmacher, F., & Kuhbandner, C. (2020). Detailed long-term memory for unattended, irrelevant, and incidentally encoded auditory information. Journal of Experimental Psychology: General, 149(2), 222–229. https:// doi.org/10.1037/xge0000650

Jonides, J. (1983). Further toward a model of the mind's eye's movement. Bulletin of the Psychonomic Society, 21(4), 247–250. https://doi.org/10.3758/BF03334699

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information.

Proceedings of the National Academy of Sciences of the United States of America, 115(43), 11084–11089. https:// doi.org/10.1073/pnas.1800006115
Katayama, J., & Polich, J. (1999). Auditory and visual P300 topography from a 3 stimulus paradigm. Clinical Neurophysiology, 110(3), 463–468. https://doi.org/10.1016/S1388-2457(98)00035-2
Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. Proceedings of the National Academy of Sciences, 111(24), 8997– 9002. https://doi.org/10.1073/pnas.1319438111
Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An activation likelihood estimation meta-analysis. Human Brain Mapping, 34(4), 814–836. https://doi.org/10.1002/hbm.21474
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
Kuhbandner, C., Rosas-Corona, E. A., & Spachtholz, P. (2017). High-fidelity visual long-term memory within an unattended blink of an eye. Frontiers in Psychology, 8. https://www.front iersin.org/articles/10.3389/fpsyg.2017.01859
LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. Psychological Review, 96(1), 101–124. https://doi.org/10.1037/0033-295X.96.1.101
LaBerge, D., Carlson, R. L., Williams, J. K., & Bunney, B. G. (1997). Shifting attention in visual space: Tests of moving-spotlight models versus an activity-distribution model. Journal of Experimental Psychology: Human Perception and Performance, 23(5), 1380–1392. https://doi.org/10.1037/0096-1523.23.5.1380
Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. Journal of Experimental Psychology: Human Perception and Performance, 21(3), 451–468. https://doi.org/ 10.1037/0096-1523.21.3.451
Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. Journal of Experimental Psychology: General, 133(3), 339–354. https:// doi.org/10.1037/0096-3445.133.3.339
Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. Journal of Experimental Psychology: Human Perception and Performance, 35(5), 1346–1358. https://doi.org/10.1037/ a0016454
Lin, J. Y., Pype, A. D., Murray, S. O., & Boynton, G. M. (2010). Enhanced memory for scenes presented at behaviorally relevant points in time. PLOS Biology, 8(3), Article e1000337. https://doi. org/10.1371/journal.pbio.1000337
Linden, D. E. J., Prvulovic, D., Formisano, E., Völlinger, M., Zanella, F. E., Goebel, R., & Dierks, T. (1999). The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. Cerebral Cortex, 9(8), 815–823. https://doi.org/10.1093/cercor/9.8.815 Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1(1), 6–21. https://doi.org/10.1080/17470214808416738

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1(1), 6–21. https://doi.org/10.1080/17470214808416738

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

Mandal, A., Liesefeld, A. M., & Liesefeld, H. R. (2022). The surprising robustness of visual attention against concurrent auditory distraction. PsyArXiv. https://doi.org/10.31234/osf.io/hnvdm

Norman, D. A. (1968). Toward a theory of memory and attention. Psychological Review, 75(6), 522–536. https://doi.org/10.1037/h0026699

Parmentier, F. B. R., Elford, G., Escera, C., Andrés, P., & San Miguel, I. (2008). The cognitive locus of distraction by acoustic novelty in the crossmodal oddball task. Cognition, 106(1), 408–432. https://doi.org/10.1016/j .cognition.2007.03.008

Piczak, K. J. (2015). ESC: Dataset for environmental sound classification (Version 2) [Dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/ YDEPUT

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32(1), 3–25. https://doi.org/10.1080/ 00335558008248231

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial Vision, 3(3), 179–197. https://doi.org/10.1163/156856888X00122

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. Science, 278(5343), 1616–1619. https://doi.org/10.1126/science.278.5343.1616

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. Science, 286(5449), 2504–2507. https://doi.org/10.1126/science.286.5449.2504

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., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. Attention, Perception, & Psychophysics, 75(3), 426–439. https://doi.org/10.3758/s13414-012-0413-x

Ruz, M., Wolmetz, M. E., Tudela, P., & McCandliss, B. D. (2005a). Two brain pathways for attended and ignored words. NeuroImage, 27(4), 852–861. https://doi.org/10.1016/j.neuroimage.2005.05.031

Ruz, M., Worden, M. S., Tudela, P., & McCandliss, B. D. (2005b). Inattentional amnesia to words in a high attentional load task. Journal of Cognitive Neuroscience, 17(5), 768–776. https://doi. org/10.1162/0898929053747685

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. Canadian Journal of Experimental Psychology/ Revue Canadienne de Psychologie Expérimentale, 66(1), 44–50. https:// doi.org/10.1037/a0025111

Smallwood, J., & Schooler, J. W. (2006). The restless mind. Psychological Bulletin, 132(6), 946–958. https://doi.org/10.1037/0033- 2909.132.6.946
Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. Proceedings of the National Academy of Sciences, 118(33), Article e2021905118. https://doi.org/10.1073/ pnas.2021905118
Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. Psychological Review, 102(3), 503–532. https://doi.org/10.1037/0033-295X.102.3.503
Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(6), 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213
Stevens, A. A., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2000). Eventrelated fMRI of auditory and visual oddball tasks. Magnetic Resonance Imaging, 18(5), 495–502. https://doi.org/10.1016/S0730-725X(00)00128-4
Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. Behavior Research Methods, 42(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096 Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology, 44(1), 24–31. https://doi.org/10.1177/0098628316677643
Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. Cognition, 115(1), 118–132. https://doi.org/10. 1016/j.cognition.2009.12.003
Terashima, H., Kihara, K., Kawahara, J. I., & Kondo, H. M. (2021). Common principles underlie the fluctuation of auditory and visual sustained attention. Quarterly Journal of Experimental Psychology, 74(4), 705–715. https://doi.org/10.1177/1747021820972255
Turk-Browne, N. B., Yi, D.-J., & Chun, M. M. (2006). Linking implicit and explicit memory: common encoding factors and shared representations. Neuron, 49(6), 917–927. https://doi.org/10.1016/j.neuron.2006.01.030
Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. Neurobiology of Learning and Memory, 91(2), 139–154. https://doi.org/10.1016/j.nlm.2008.10. 011
Vecera, S. P., & Behrmann, M. (2001). Attention and unit formation: A biased competition account of object-based attention. In Advances in psychology (Vol. 130, pp. 145–180). Elsevier.
von Restorff, Hedwig (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. Psychologische Forschung [Psychological Research], 18(1), 299– 342. (in German). https://doi.org/10.1007/BF02409636. S2CID 145479042
Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. Science, 281(5380), 1188–1191. https://doi.org/10.1126/science .281.5380.1188
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, Article 105201. https://doi.org/10 .1016/j.cognition.2022.105201

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. Psychological Bulletin, 63(6), 410–424. https://doi.org/10.1037/h0022001

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. https:// doi.org/10.1037/0022-3514.54.6.1063

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

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 14883. https://doi.org/10 .1038/s41598-021-94161-0

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. Nature Neuroscience, 7(9), Article 9. https:// doi.org/10.1038/nn1294

# Chapter 5: Engagement strengthens neural representations of naturalistic narratives

The field of cognitive neuroscience has seen increased enthusiasm in naturalistic experimental contexts with the aim of testing cognitive and neural mechanisms under contexts more aligned with real life experience (Natase et al., 2020; Sonkusare et al., 2019). Naturalistic task paradigms such as movie watching and story listening benefit by being minimally controlled, allowing participants to experience them nearly as they would if they were watching a movie or listening to a story outside of a laboratory setting. For example, as we comprehend a narrative, our subjective engagement changes as we experience relative increases and decreases in interest, immersion, and attentional capture by the story. In the current study, we leverage natural viewing and listening experiences to investigate the neural correlates of engagement fluctuations under unconstrained conditions as would be experienced in daily life.

Subjective engagement indexes the level of engrossment or immersion an individual experiences as they comprehend a story over time (Busselle & Bilandzic, 2009). Momentary engagement may be driven by a confluence of factors, including attentional state, arousal levels, emotional valence, personal relevance, and salience of a stimulus. Indeed, these features may interact to result in one's experience of engagement while watching movies or listening to stories.

Fluctuations in subjective engagement dynamically impact patterns of fMRI neural activity and connectivity. Distributed functional networks reliably reconfigure to index momentary engagement during narrative viewing and listening (Corriveau et al., 2026; Song et al., 2021). Individuals who experience narratives more similarly show increased neural synchrony in distributed brain regions, including the default mode network (Nanni-Zepeda et al., 2024; Ohad & Yeshurun, 2023). Moreover, analog measures provide insight into neural mechanisms involved in

engagement. Fluctuations in attentional state are accompanied by dynamic changes in distributed neural activity (Esterman et al., 2013; Kucyi et al., 2017) and functional networks (Jones et al, 2024; Rosenberg et al., 2020; Zuberer et al., 2021), which additionally generalize to predict subjective engagement (Corriveau et al., 2026; Song et al., 2021). Relatedly, mind-wandering—a proxy for task *dis*engagement—is reflected in network activity and connectivity, especially related to the default mode network (Christoff et al., 2016; Kucyi, 2018). Arousal fluctuations, likely tightly-coupled to engagement dynamics during narratives, are related to neural synchrony across individuals (Nummenmaa et al., 2012; 2014) and dynamically reconfigure functional brain networks during ongoing narratives (Ke et al., 2025; Park et al., 2026; Young et al., 2017) suggesting widespread neural changes that track fluctuating internal states.

Less understood is the extent to which fluctuations in ongoing engagement dynamically impact the content of neural representations. Said another way, when we experience heightened engagement during a movie or story, is the information from that narrative more strongly reflected in neural activity patterns? Previous work suggests that this may be the case. Selective attention strengthens representations of relevant features across cortex (Erez & Duncan, 2015; Peelen et al., 2009; Sprague & Serences, 2013) and in the hippocampus (Aly & Turk-Browne, 2016) and shifts neural tuning toward relevant stimulus information (Brouwer & Heeger, 2013; Cukur et al., 2013; Nastase et al., 2017). Moment-by-moment changes in attentional state modulate neural alignment across subjects, suggesting individuals more similarly and precisely represent stimulus content under heightened attentional states (Rothlein et al., 2018) and during key moments in a narrative (Upadhyahula et al., 2025). However, beyond more similar patterns of neural activity, it is not clear whether fluctuations in attention reflect a convergence of neural activity around a shared mental state—i.e., similar neural patterns reflect similar activity in response to increases in

attentional capture, engagement, suspense, etc.,—or whether alignment of neural activity reflects a shift towards a shared representation of the narrative stimulus content.

Until relatively recently, the ability to model the content of narrative stimuli has been burdensome, requiring countless hours of experimenter-provided (and often research assistant-provided) subjective ratings of stimulus attributes or explicit controlled manipulation of key features during task design. However, the development of neural networks trained to represent stimulus features (e.g., the identity of an object depicted in an image) provides an invaluable resource for modelling relevant stimulus content. In the current study, we utilize convolutional neural networks trained on visual image recognition (AlexNet; Krizhevsky et al., 2012) and speech recognition (wav2vec 2.0; Baevski et al., 2020) to obtain vectorized representations of visual and auditory content. Through use of these models, we are able to derive "ground truth" representations of stimulus content which can then be compared to neural activity. Using a representational similarity analysis framework (Kriegeskorte et al., 2008), we can compare whether patterns of brain activity better reflect the relationships between unique scenes within the same neural-network modeled representational space under moments of heightened engagement, as would be predicted by past work (Rothlein et al., 2018).

Further, these convolutional network architectures are hierarchical such that earlier model layers perform computations that evaluate low-level visual and auditory information, whereas later layers integrate this information into high-level classification to perform word and object recognition (Jozwik et al., 2017; Pasad et al., 2022). Investigation of AlexNet demonstrates that layer representations correspond well with the hierarchy of the visual system (Güçlü & van Gerven, 2015; Wagatsuma et al., 2022). As such, neural networks can be used to derive objective

representations of relatively low- and high-level information to test whether engagement differentially impacts alignment with neural activity.

In the current study, we examine whether dynamic fluctuations in subjective engagement during narratives impacts the representational fidelity of stimulus information in the brain. Using convolutional neural networks, we model visual and auditory information experienced by participants as they comprehend a narrative. Using representational similarity analyses, we then compare whether neural representations throughout the brain are more closely aligned with modeled representations during periods of heightened engagement, as would be predicted if engagement unites neural activity towards a common, stimulus-driven representational space. Results reveal that engagement increases representational alignment of stimulus content throughout the brain, suggesting widespread impacts of engagement on the fidelity of stimulus information in neural signals.

## Methods

### *fMRI naturalistic narratives*

Participants (N=60) completed a two-session fMRI experiment, approximately one week apart (mean days between sessions=10.88, SD=9.87). Data was collected on a 3-Tesla Philips Ingenia scanner. Voxels were 2.526mm x 2.526mm x 3mm. Volumes were collected using a multiband sequence with a repetition time of 1 second. All study procedures were approved by the Social and Behavioral Science Institutional Review Board at the University of Chicago.

Over two sessions, participants passively viewed four naturalistic narrative stimuli in the scanner–an auditory-only podcast (*moviePie*, 16 min 34 s) in which a baker describes the process of baking a berry pie from scratch, a visual-only silent movie (*movieCross*, 16 min 59 s) which

shows the process of baking croissants with strawberries and cream, an audio-visual movie (*movieCake*, 15 min 40 s) in which a baker shows and describes the process of making a tiered chocolate cherry cake, and an audio-visual clip (*movieNorth*, 14 min 49 s) from the movie North by Northwest featuring dialogue between two characters and a suspenseful airplane chase scene. The order of stimulus presentation was quasi-counterbalanced across fMRI sessions such that participants saw one single-modality stimulus (audio-only or visual-only) and one audio-visual stimulus per scan. The single-modality stimulus was always presented first, followed by the audio-visual stimulus. Following both narratives, participants performed an annotated rest task (10 min) and finally, a controlled sustained attention task. Rest and sustained attention tasks are not analyzed in the current study.

*Engagement ratings*

Following both scans, participants completed a subjective engagement rating task for the narrative stimuli presented during the fMRI scan. Participants saw or heard each narrative stimulus a second time and were instructed to continuously indicate their subjective level of engagement throughout its entire duration using a slider bar on a scale ranging from "Not engaging at all" to "Completely engaging." The scale contained 25 steps and always initialized in the middle (on step 13). Participants pressed the arrow keys on a keyboard to report small changes in engagement and could indicate large changes (5 steps) in engagement by pressing the shift key and an arrow key simultaneously. Prior to the subjective engagement rating task, participants also completed a series of memory tasks for both the naturalistic and controlled sustained attention task stimuli. Memory data are not analyzed in the current study.

We averaged individual engagement time series to create a group-average engagement time series. Previous work has shown that the group-average time course captures stimulus-driven engagement, where peaks and troughs correspond to sensical highly-engaging (suspenseful, surprising) and minimally-engaging (repetitive, boring) moments in a narrative (Song et al., 2021). Averaging engagement across a group also removes possible idiosyncratic differences in engagement between first (in-scanner) and second viewings. We included time series from participants who reported a change in their subjective engagement at least once per minute on average, indicating task compliance. This criterion resulted in the exclusion of 3 time series from *moviePie*, 3 time series from *movieCross*, 3 time series from *movieCake*, and 3 time series from *movieNorth*. Average engagement time series therefore included the following sample sizes: *moviePie N*=54, *movieCross N*=55, *movieCake N*=50, *movieNorth N*=54.

*fMRI preprocessing*

Functional MRI data were preprocessed using Analysis of Functional Neuroimages (AFNI version 19.0). The first 3 TRs of each functional run were excluded from analysis. Despiking, slice-time correction, and motion-correction was applied to the data. We regressed covariates of no interest from the data, including a 24-parameter head motion model (6 motion parameters, 6 temporal derivatives, and their squares) and mean signal from subject-specific eroded white matter and ventricle masks. We then registered the functional images to participants' skull-stripped MPRAGE anatomical images with linear transformation and normalized to the Montreal Neurological Institute (MNI) space with nonlinear warping.

*Visual modeled representations*

The content of visual narratives was modeled at the level of scenes. Individual scenes were identified using the Python package PySceneDetect (adaptive detector threshold=4.0, minimum scene length=1s) which automatically detects changes in visual information to parse videos into individual scenes. Scenes were required to be at least 1 second in length. Narrative stimuli were segmented into variable numbers of visual scenes: 229 scenes in *movieCross*, 158 scenes in *movieCake*, 173 scenes in *movieNorth*.

We sought to model the average low-level and high-level visual properties of each scene. To do so, we utilized the pretrained convolutional neural network (CNN) AlexNet (Krizhevsky et al., 2012). AlexNet is a network trained for image recognition. Its architecture consists of five convolutional layers (some of which are followed by max-pooling layers) and three fully-connected layers. Layer output was extracted using the *Thingsvision* package in python (Muttenhaler & Hebart, 2021). Frames were passed through the pretrained AlexNet network in the *torchvision* library. We extracted output from each of the five convolutional layers and three fully-connected layers. We then averaged the output for all frames in a scene. For main analyses, we compare representations obtained from layer 1 and layer 8, whose output has previously been shown to correspond well with neural activity occurring in low-level visual (V1) and higher-order visual regions, respectively (Güçlü & van Gerven, 2015; Wagatsuma et al, 2022).

*Auditory modeled representations*

We next modeled auditory content from narratives containing auditory information, i.e., the audio-visual movies and podcast. The spoken content of narratives was transcribed and time-stamped using the automatic speech recognition system Whisper (Radford et al., 2022) and

transcription errors were manually corrected. Whisper automatically segmented auditory content into discrete units which were most often single sentences but were sometimes shorter (individual phrases) and other times longer (multiple sentences). For simplicity, we used the segmentations and timings (start and end of segmented speech) provided by Whisper. Narrative stimuli were divided into the following number of auditory segments: 193 segments in *moviePie*, 245 segments in *movieNorth*, 133 segments in *movieCake*.

To model auditory features, we extracted output from the pretrained base wav2vec 2.0 model from the *torchaudio* library, a convolutional neural network trained on speech recognition. Its architecture includes a 7-layer convolutional network feature encoder followed by 12 transformer layers. Layer output was extracted for each auditory segment in a narrative. For main analyses, we compared representations extracted from transformer layer 1 whose output reflects local acoustic features, and layer 8 whose output contains information about word identity and meaning (Pasad et al., 2022). Results from all transformer layers are included in supplementary materials.

*Neural activations - Least Squares Separate*

For each visual scene or auditory segment, we fit fMRI data with a least-squares separate general linear models (LSS GLMs; Mumford et al., 2014) in which each scene was modeled in a separate design matrix as a regressor of interest, with all other scenes included as a combined nuisance regressor. Regressors were convolved with a canonical Glover HRF and its temporal and dispersion derivatives. We extracted voxel-wise T-statistics which reflect noise-normalized patterns of activation for each scene. For representational similarity analysis, voxels were grouped into 400 cortical regions (Schaefer et al., 2018) and 32 subcortical regions of interest (Tian et al.,

2022). For display purposes, regions are also grouped into canonical resting-state networks (Yeo et al., 2011).

*Representational similarity analysis*

In the current study, our primary aim was to test whether engagement strengthened stimulus representations in the brain. To this end, we sought to model the content of narrative stimuli in a manner that enabled us to compare model-derived representations with patterns of neural activity. This is possible with representational similarity analysis (RSA; Kriegeskorte et al., 2008) which evaluates shared representational structure within a common dissimilarity space. Briefly, RSA calculates a measure of dissimilarity (e.g., Euclidean distance) between pairs of representational vectors—for example, dissimilarity in CNN layer 1 output for all scenes in a movie—resulting in a representational dissimilarity matrix (RDM). Dissimilarity matrices for two different representational spaces can then be directly compared using a metric such as Pearson's correlation.

In the current study, we were interested in the alignment between neural and model-derived RDMs as a function of subjective engagement. Therefore, we calculated separate RDMs for scenes that occurred during moments of low- or high-engagement. Mean engagement time courses were divided into thirds. High-engagement scenes were those in which the majority of TRs were categorized as high-engagement. Low-engagement scenes were scenes in which the majority of TRs in the scene fell within a low-engagement period.

Dissimilarity between voxel-wise t-statistics within regions of interest was calculated using Euclidean distance. This resulted in one high-engagement and one low-engagement RDM per region. Dissimilarity between model-derived (CNN) representational vectors was calculated as 1

minus the cosine similarity between vectors. This resulted in one high-engagement and one low-engagement RDM per metric. The alignment between model-derived and neural RDMs was computed using the Fisher-z normalized partial Pearson correlation controlling for similarity in the onset time of the scene or segment. Finally, to compare alignment between model-derived and neural representations during moments of high and low engagement, we conducted a paired t-test of the alignment scores between high- and low-engagement RDMs for each brain region. False discovery rate was controlled with the Benjamini-Hochberg p-value adjustment, $q<.10$. Methods are depicted graphically in **Figure 25**.

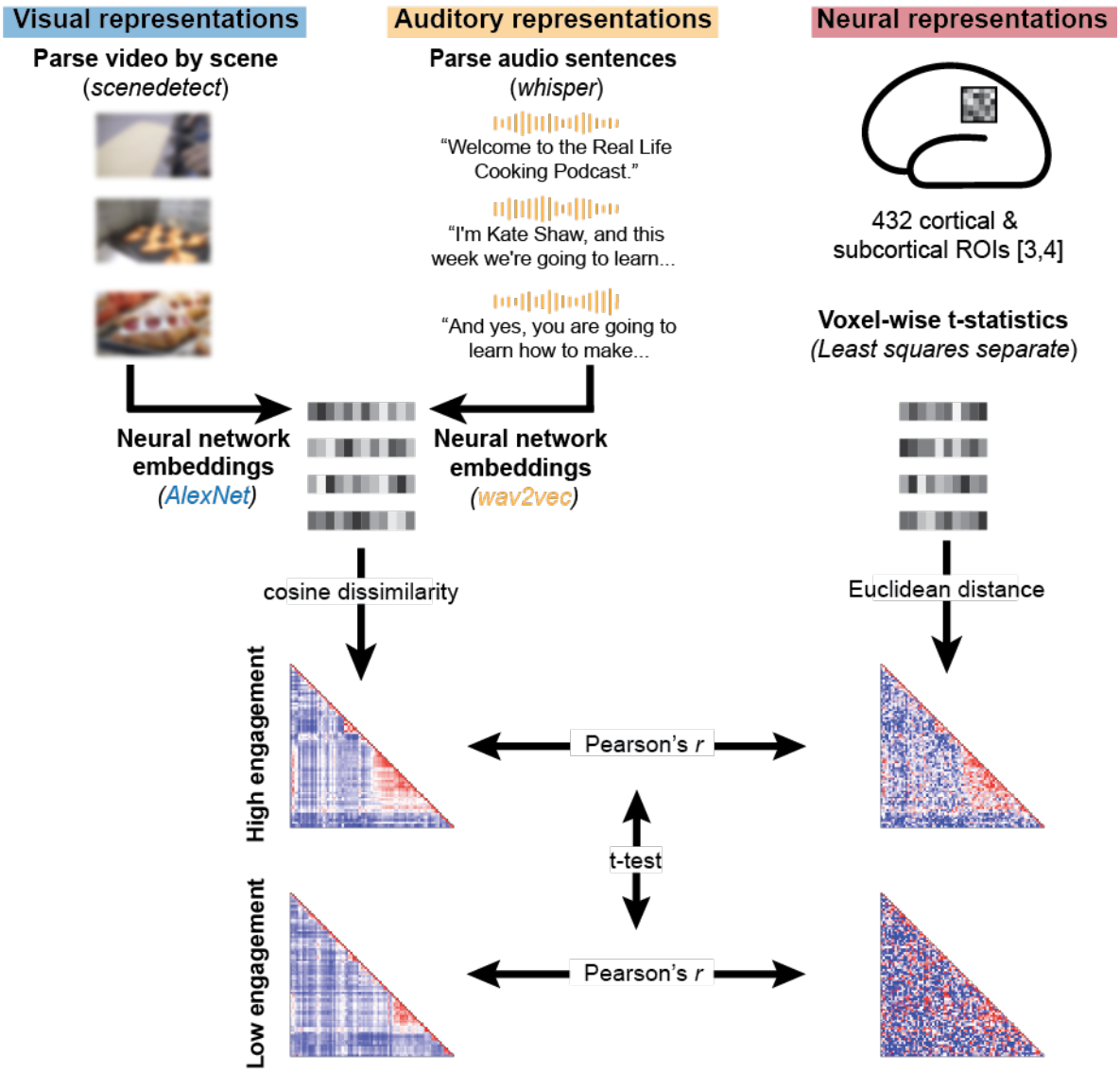



*Figure 25. Representational similarity methods.* To model the representational space of narrative content, movies with a visual component were segmented into discrete, visually-similar scenes using the *scenedetect* package implemented in Python. Narratives with an auditory component were parsed into discrete auditory segments using the *Whisper* package in Python. We extracted layer output from neural networks (*AlexNet* and *wav2vec2.0*) comprising vectorized representations of visual and auditory narrative content. Least-squares separate general linear models provided voxel-wise t-statistics within each of 432 cortical and subcortical ROIs. Using a

tertile split, scene- and auditory-segments were categorized as either high- or low-engagement and representational dissimilarity matrices (RDMs) were calculated using either cosine dissimilarity (neural-network representations) or Euclidean distance (brain representations). Finally, we compared alignment (Pearson's *r*) between high- and low-engagement RDMs using a paired t-test.

**Results**

*Engagement strengthens high-level representational alignment*

We first tested whether representational alignment of late-layer—and therefore high-level—neural network representations are strengthened under periods of heightened subjective engagement. We performed representational similarity analyses between neural network-derived visual (AlexNet) and auditory (wav2vec 2.0) representations for narrative segments that occurred during periods of high vs. low engagement. We hypothesized that heightened states of engagement would result in more faithful representations of stimulus content as compared to periods of low engagement. Results for all narratives are displayed in **Figure 26**.

In all three movies with visual content, we observed that similarity between neural network-modeled representations and patterns of neural activity is stronger in distributed regions during periods of increased engagement (**Figure 26A**). The individual regions for which this is true varies by stimulus. When comparing regions that survived FDR-correction in all three movies, we see a relatively large proportion of somatomotor, salience/ventral attention, and subcortical networks whose activity patterns are more aligned with neural network-modeled representations during heightened engagement. Consistent modulation in alignment between neural and network-modeled visual representations suggests that periods of higher engagement may shift patterns of neural activity towards a more objective representational space of stimulus content.

Similarly, in all three narratives containing spoken content, representational alignment between neural network-modeled representations and neural activity patterns is generally increased under periods of better engagement, although the regions that survive correction are again distributed by stimulus (**Figure 26B**). Relatively fewer regions showed this shift for auditory vs. visual representational alignment, with the highest proportion of regions coming from the dorsal attention network. However, general increases in alignment across auditory narratives further support the possibility that engagement results in higher agreement of participants' neural activity in representing auditory stimulus features.

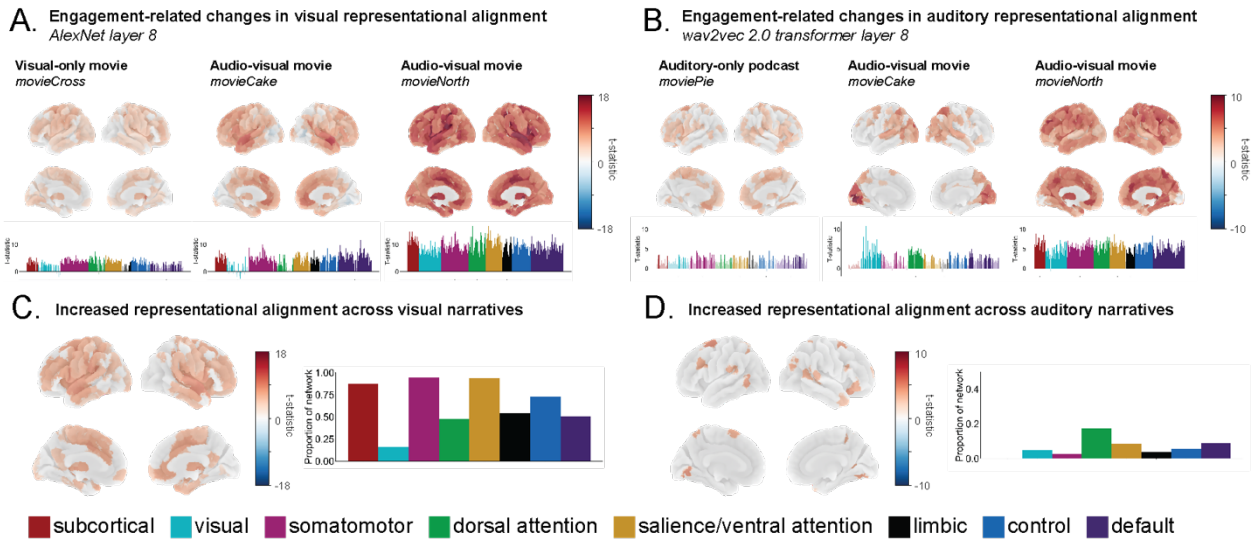



*Figure 26. Representational alignment between brain and modeled representations strengthens with engagement.* Alignment between neural activity and (A) visual and (B) auditory representations as modeled with convolutional neural networks is strengthened under heightened engagement across participants. Colored brain regions reflect regions whose representational alignment is stronger under high (red) and low (blue) engagement, FDR-corrected $q$<.10. The full range of t-statistics from the paired t-test across participants is plotted beneath brain figures, sorted

by canonical resting-state networks (Yeo et al., 2011; Tian et al., 2022). Regions not surviving FDR-correction are shown with decreased opacity. For visualization purposes, regions that survived FDR correction across narratives are plotted in the second row (C-D), as well as the relative proportion of regions within each canonical resting-state network that showed increased alignment across narratives. Brain regions reflect the average t-statistic across narratives.

*Low-level representational alignment is strengthened under engagement*

A key feature of the neural networks used in the current study is that their hierarchical architecture provides representations that align with low-level as well as high-level properties useful for image and speech recognition. As such, we can examine whether alignment between neural patterns and representations obtained from an early network layer—reflecting low-level visual and auditory properties—is similarly increased under moments of higher engagement. We hypothesized that representations of low-level stimulus features, especially those in sensory areas, would be less impacted by engagement as the processing of low-level information in these areas is more reliable overall (Hasson et al., 2010) and may therefore be robust to modulation by attentional engagement.

Contrary to our hypothesis, we found that alignment between neural activity and low-level representations was largely strengthened under attention (**Figure 27**). This was particularly true for both visual and auditory low-level representations during audio-visual movies. Few regions showed reliable changes in alignment of low-level visual information, with the largest contribution of dorsal attention network regions. We did not observe any regions whose representational alignment for low-level auditory information was reliably impacted by engagement across narratives. Overall, however, general increases in representational alignment across movies

suggest that engagement impacts patterns of neural activity related to low-level stimulus information such that their representational space is more similar to that which would be predicted by a neural network when more engaged.

One possibility for similar changes in alignment between earlier and later neural network layers may be that low- and high-level properties are correlated with one another. In other words, visual scenes with similar high-level content (e.g., containing the same objects) may also share more low-level features (e.g., color), resulting in correlated representations in dissimilarity space. To test this possibility, we correlated the RDMs calculated from early and late neural network outputs. Visual RDMs calculated between all pairs of visual scenes were positively correlated across early and late AlexNet layers (layers 1 and 8) in all movies (*movieCross*: $r$=0.580; *movieCake*: $r$=0.565; *movieNorth*: $r$=0.580; all $p$s<.001). Similarly, RDMs of all auditory segments calculated from early and late layers of wav2vec 2.0 (transformer layers 1 and 8) were positively correlated in all auditory narratives (*moviePie*: $r$=0.834; *movieCake*: $r$=0.847; *movieNorth*: $r$=0.389, all $p$s<.001). This suggests that high- and low-level features were indeed related, making similar modulation by engagement less surprising.

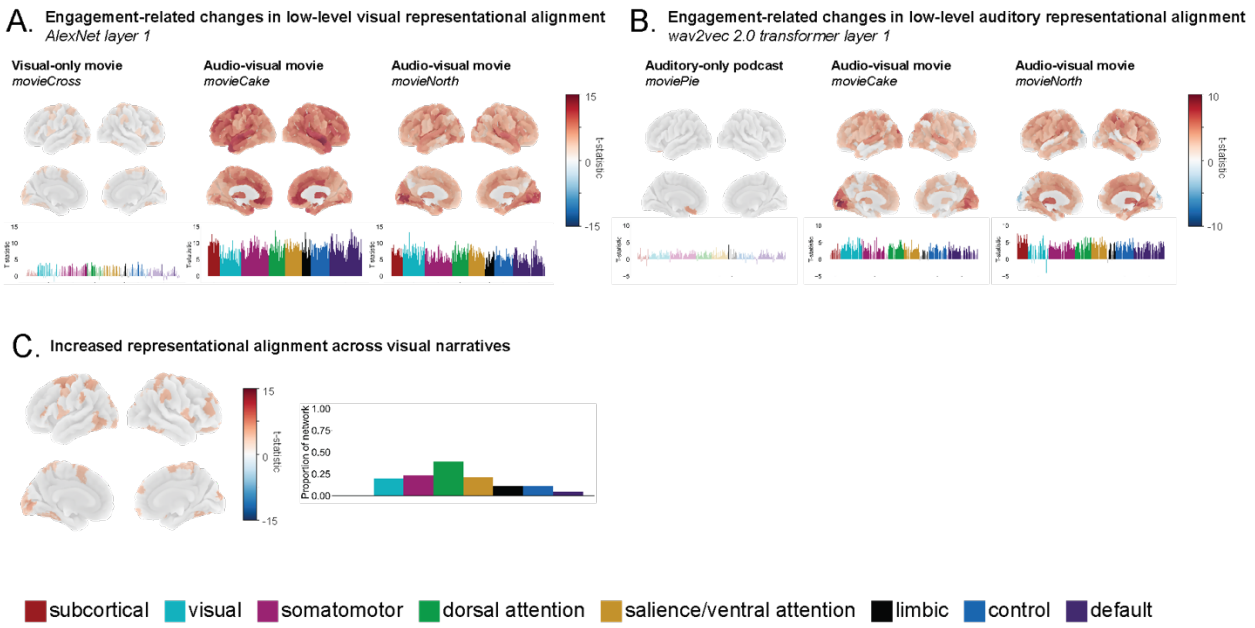



*Figure 27. Engagement strengthens representational alignment between neural and low-level representations.* Alignment between low-level, i.e., early network layer, (A) visual and (B) auditory representations and patterns of neural activity are strengthened during audio-visual movies. Colored brain regions reflect regions whose representational alignment is stronger under high (red) and low (blue) engagement, FDR-corrected $q$<.10. The full range of t-statistics from the paired t-test across participants is plotted beneath brain figures, sorted by canonical resting-state networks (Yeo et al., 2011; Tian et al., 2022). Regions not surviving FDR-correction are shown with decreased opacity. (C) Regions whose alignment with visual representations was strengthened under high engagement across narratives are plotted in the second row, as well as the relative proportion of regions within each canonical resting-state network that showed increased alignment across narratives. Brain regions reflect the average t-statistic across narratives. No regions were reliably more aligned with auditory representations under better engagement across narratives.

## Discussion

The primary aim of the current study was to investigate the impacts of naturally-fluctuating attentional engagement on the fidelity of stimulus representations. Using representations of visual and auditory stimulus features obtained from neural networks, we modeled the extent to which engagement shifted neural representations towards a more objective similarity space. We hypothesized that patterns of neural activity would align more with neural-network-modeled representations during periods of higher vs. lower attentional engagement. Indeed, this is what we observed—activity in regions distributed throughout the brain were more aligned with modeled representations when participants subjectively reported being more engaged with the narrative, suggesting neural activity patterns may be more stimulus-driven during these times.

Previous work demonstrates an increase in neural synchrony observed during moments of heightened engagement (Nanni-Zepeda et al., 2024; Ohad & Yeshurun, 2023), sustained attentional states (Rothlein et al., 2018), arousal (Nummenma et al., 2012; 2014), and key moments (Upadhyayula et al., 2025). The current work extends these findings, providing a possible representational space within which neural activity patterns align. The convergence of neural activity observed across participants under similar mental states, rather than reflecting the mental state itself, may reflect stronger agreement in objective representations of stimulus content. The use of neural networks in the current study to model stimulus content itself substantiates this possibility by providing a "ground-truth" representational space to which neural activity can be compared. In this way, neural networks can greatly lower the experimenter burden of manually annotating stimuli in naturalistic designs by providing an objective quantification of stimulus information.

Contrary to our expectations, engagement strengthened representational alignment for low-level visual and auditory information as well as high-level information throughout the brain. This was not initially hypothesized, as we predicted that low-level features would be represented more strongly in sensory regions (Güçlü & van Gerven, 2015; Wagatsuma et al., 2022) whose activity is more reliable during narratives (Hasson et al., 2010). Therefore, we did not expect engagement to strongly modulate low-level representations, particularly in visual and auditory cortices. However, positive correlations between RDMs calculated from early and late network layers suggest that low- and high-level features were strongly related in the narratives, suggesting that engagement should similarly impact alignment of neural patterns to both RDMs. Future work may seek to dissociate low- and high-level stimulus content to better determine whether neural alignment to these representational spaces is differentially modulated by engagement.

In the current study, we chose to model visual and auditory information as a straightforward means of capturing stimulus content with which we could compare neural data. It was also convenient; many neural networks like those used in this study have been trained on tasks related to computer vision and speech recognition and are freely available online. We do not expect our results to be limited to the neural networks selected for this study. Rather, we expect the representational dissimilarity spaces to be largely similar across neural networks trained to perform similar tasks. Further, the continuous advancement and development of convolutional neural networks to perform a variety of tasks provides a means of modeling endless stimulus content. Future work may seek to test the extent to which representation of other stimulus features (e.g., the memorability of scenes; Needell & Bainbridge, 2022) is impacted by engagement. Additionally, visual and auditory stimulus representations in the current study were modeled discretely, such that their representations did not include any information about the context in

which they appeared. True representations may change over time and be influenced by the broader context of the narrative in a way that is not captured by the current methods. Methods that consider these variable representations may be interesting directions for future research.

## Conclusion

As we naturalistically perceive a stimulus, our experience of engagement dynamically fluctuates over time. Using convolutional neural networks, we demonstrate that our subjective experience of engagement corresponds with marked shifts in stimulus-relevant representations in patterns of neural activity throughout the brain. These results suggest that periods of heightened engrossment in a narrative may lead neural activity to be more stimulus-driven, such that representations are more closely aligned within a network-modeled representational space. Findings complement a growing body of work seeking to characterize shared neural responses to naturalistic narratives and provide a methodological approach for testing hypotheses about the content of neural representations.

**Chapter 5 Bibliography**


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

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20*, 12449–12460. https://dl.acm.org/doi/10.5555/3495724.3496768

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience*, *33*(39), 15454–15465. https://doi.org/10.1523/JNEUROSCI.2472-13.2013

Busselle, R., & Bilandzic, H. (2009). Measuring Narrative Engagement. *Media Psychology*, *12*(4), 321–347. https://doi.org/10.1080/15213260903287259

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

Corriveau, A., Ke, J., & Rosenberg, M. D. (2026). Common Brain Network Dynamics Capture Attention Fluctuations in Tasks and Movies. *Journal of Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/JOCN.a.2622

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *Journal of Neuroscience*, *35*(36), 12383–12393. https://doi.org/10.1523/JNEUROSCI.1134-15.2015

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

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Jones, H. M., Yoo, K., Chun, M. M., & Rosenberg, M. D. (2024). Edge-Based General Linear Models Capture Moment-to-Moment Fluctuations in Attention. *Journal of Neuroscience*, *44*(14). https://doi.org/10.1523/JNEUROSCI.1543-23.2024

Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., & Mur, M. (2017). Deep Convolutional Neural Networks Outperform Feature-Based But Not Categorical Models in Explaining Object Similarity Judgments. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01726

Ke, J., Song, H., Bai, Z., Rosenberg, M. D., & Leong, Y. C. (2025). Dynamic brain connectivity predicts emotional arousal during naturalistic movie-watching. *PLOS Computational Biology*, *21*(4), e1012994. https://doi.org/10.1371/journal.pcbi.1012994

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*,

*25*. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Kucyi, A. (2018). Just a thought: How mind-wandering is represented in dynamic brain connectivity. *NeuroImage, Brain Connectivity Dynamics*, *180*, 505–514. https://doi.org/10.1016/j.neuroimage.2017.07.001

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

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Nanni-Zepeda, M., DeGutis, J., Wu, C., Rothlein, D., Fan, Y., Grimm, S., Walter, M., Esterman, M., & Zuberer, A. (2024). Neural signatures of shared subjective affective engagement and disengagement during movie viewing. *Human Brain Mapping*, *45*(4), e26622. https://doi.org/10.1002/hbm.26622

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. *Cerebral Cortex (New York, NY)*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Needell, C. D., & Bainbridge, W. A. (2022). Embracing New Techniques in Deep Learning for Estimating Image Memorability. *Computational Brain & Behavior*, *5*(2), 168–184. https://doi.org/10.1007/s42113-022-00126-5

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, *109*(24), 9599–9604. https://doi.org/10.1073/pnas.1206095109

Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jääskeläinen, I. P., Hari, R., & Sams, M. (2014). Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *NeuroImage*, *102*, 498–509. https://doi.org/10.1016/j.neuroimage.2014.07.063

Ohad, T., & Yeshurun, Y. (2023). Neural synchronization as a function of engagement with the narrative. *NeuroImage*, *276*, 120215. https://doi.org/10.1016/j.neuroimage.2023.120215

Park, J. S., Gollapudi, K., Ke, J., Nau, M., Pappas, I., & Leong, Y. C. (2026). Emotional arousal enhances narrative memories through functional integration of large-scale brain networks. *Nature Human Behaviour*, *10*(2), 370–383. https://doi.org/10.1038/s41562-025-02315-1

Pasad, A., Chou, J.-C., & Livescu, K. (2021). Layer-Wise Analysis of a Self-Supervised Speech Representation Model. *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 914–921. https://doi.org/10.1109/ASRU51503.2021.9688093

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning, ICML'23*, *202*, 28492–28518.

Rosenberg, M. D., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., Ramani, R., Qiu, M., Constable, R. T., & Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, *117*(7), 3797–3807. https://doi.org/10.1073/pnas.1912226117

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

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proceedings of the National Academy of Sciences*, *118*(33), e2021905118. https://doi.org/10.1073/pnas.2021905118

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. https://doi.org/10.1038/nn.3574

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, *23*(11), 1421–1432. https://doi.org/10.1038/s41593-020-00711-6

Upadhyayula, A., Henderson, J. M., Zacks, J. M., & Reagh, Z. M. (2025). *Key moments in naturalistic events synchronize neural activity patterns and dominate memory reinstatement* (p. 2025.08.30.673233). bioRxiv. https://doi.org/10.1101/2025.08.30.673233

Young, C. B., Raz, G., Everaerd, D., Beckmann, C. F., Tendolkar, I., Hendler, T., Fernández, G., & Hermans, E. J. (2017). Dynamic Shifts in Large-Scale Brain Network Balance As a Function of Arousal. *Journal of Neuroscience*, *37*(2), 281–290. https://doi.org/10.1523/JNEUROSCI.1759-16.2016

Zuberer, A., Kucyi, A., Yamashita, A., Wu, C. M., Walter, M., Valera, E. M., & Esterman, M. (2021). Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *NeuroImage*, *229*, 117610. https://doi.org/10.1016/j.neuroimage.2020.117610

# Chapter 6: Sustained and selective attention jointly impact neural decoding of visual stimuli

Despite our best intentions, humans struggle to maintain our focus of attention on task-relevant goals for extended periods of time, instead fluctuating between periods of in-the-zone and error-prone, out-of-the-zone processing. While this experience is ubiquitous, we lack an understanding of the impacts of sustained attentional fluctuations on the fate of task-relevant neural information. Even less certain is the impact of these fluctuations on the neural representations of information outside our task goals. The primary aim of the current study is to understand the dual influence of selective and sustained attention on the representation of visual information in the brain through investigation of EEG decoding time courses. We further extend these findings to better understand how these distinct attentional processes influence recognition memory judgements.

Previous work demonstrates that selective attention impacts how information is represented neurally. The presence of relevant visual stimuli during visual search results in stronger fMRI representations of relevant image categories (Cohen & Tong, 2015; Peelen et al., 2009), relevant features (Jackson et al., 2017; Jehee et al., 2011), and stronger decoding from fMRI signals (Erez & Duncan, 2015) than observed for irrelevant distractors. Changing task-relevant goals leads to widespread shifts in neural geometry and cortical tuning (Bracci et al., 2017; Brouwer & Heeger, 2013; Cukur et al., 2013; Harel et al,. 2014; Nastase et al., 2017) as well as patterns of hippocampal activity (Aly & Turk-Browne, 2016), suggesting that neural activity patterns adapt to accommodate attentional selection. Neural changes are also reflected in temporally-resolved time courses of visual category decoding as measured with EEG. In seminal work by Kaiser et al., (2016), searching for a relevant-category item in a scene resulted in faster

and more accurate decoding of object categories when the scene indeed contained the relevant object vs. when it contained a distractor object. In combination, selective attention strengthens representations of relevant stimuli and stimulus features, increasing the availability of this information in neural signals.

Relatively less is known about the impacts of momentary sustained attentional fluctuations, which occur over a matter of seconds (Esterman et al., 2013), on neural representations. Dynamic fluctuations in sustained attention to visual stimuli correspond to increased univariate activity in large-scale brain networks (Esterman et al., 2013; Fortenbaugh et al., 2018; Kucyi et al., 2017) and rapidly reconfigure functional connections between brain regions (Corriveau et al., 2026; Jones et al., 2024; Rosenberg et al., 2020; Zuberer et al., 2021). It is not clear whether these attentional state-related differences are impacting the neural content of stimulus-related representations, although preliminary evidence points to this possibility. Esterman et al., (2014) found greater repetition attenuation for repeated irrelevant background scenes in scene-specific cortex under better attentional states, suggesting that irrelevant scenes were processed more during these periods of heightened sustained attention. Importantly, this finding also suggests a dissociation between attentional selection and sustained attention fluctuations, such that better sustained attentional states are not accompanied by stronger filtering of task-irrelevant information. Further, patterns of EEG (Chidharom et al., 2025) and fMRI activity (deBettencourt et al., 2015; Rosenberg et al., 2015; Song et al., 2023; Yamashita et al., 2021) distinguish good vs. poor sustained attentional states, indicating that stimulus representations may reliably change as a function of sustained attention. Finally, additional evidence of sustained-attention-related representational changes come from the finding that voxel-wise patterns of fMRI activity are more aligned across

participants under better sustained attentional and motivational states (Rothlein et al., 2018), with neural alignment possibly indicating stronger representation of task-related information.

In addition to changes in processing, both selective and sustained attention impact how stimuli are subsequently remembered (Corriveau et al., 2026; Sherman & Turk-Browne, 2024). Task-relevant stimulus features are better remembered (Aly & Turk-Browne, 2016; Corriveau et al., 2025; Corriveau, Chao et al., 2024; Uncapher & Rugg, 2009), as are stimuli presented in selected locations in space (Turk-Browne et al., 2013; Uncapher et al., 2011). Information presented during better sustained attentional states, as indexed by online behavioral measures, also show superior memory performance (deBettencourt et al., 2018; Decker et al., 2023; Wakeland-Hart et al., 2022). Interestingly, the impacts of selective and sustained attention on recognition memory performance appear to be dissociable (Corriveau, Chao et al., 2024; Corriveau et al., 2025; deBettencourt et al., 2021) such that both types of attention uniquely contribute to successful memory judgements. For example, Corriveau & Chao et al., (2024) found that, atop a robust effect of selective attention (task-relevance) on memory judgements, both relevant and irrelevant images presented during better sustained attentional states were better-remembered. These results, in combination with related work (Corriveau et al., 2025), contradict the intuition of an inherent tradeoff between processing for relevant and irrelevant information and necessitate investigation into the unique impacts of selective and sustained attention on mnemonic processes.

The primary aim of the present study was to examine the dual impacts of selective and sustained attention on neural representations of visual stimuli, as indexed by EEG decoding accuracy. Participants completed a continuous performance task during which behavior provided a continuous readout of sustained attentional state while selectively attending specific stimuli within a spatially-consistent visual display. We examined how the availability of stimulus

information in neural signals varied as a function of selectivity (task-relevance) and sustained attentional state by testing EEG decoding performance under different attentional conditions. We find that, in addition to robust differences in classification as a function of selective attention (task-relevance), sustained attentional states further modulate the time course of decoding for both relevant and irrelevant stimuli. Additionally, we find that targets presented during heightened sustained attentional states, as measured by the strength of EEG decoding, are better-remembered.

## Methods

Data was collected as part of a preregistered study (https://osf.io/h9r4e/overview). Deviations from preregistered analyses are noted at the end of the Methods section. Participants were recruited from the University of Chicago and the surrounding community and were compensated for participation with cash, class credit, or a combination of cash and class credit. All aspects of the study were conducted in accordance with protocols approved by the Institutional Review Board at the University of Chicago.

The full experimental procedure included three tasks in total. Task order was always as follows: EEG localizer task, continuous performance task, memory task. Each task is described in detail below.

### *EEG localizer task*

Participants completed six blocks of the EEG localizer task, named for the fMRI tasks after which it was modeled. Participants were tasked with detecting when the fixation dot presented in the center of the screen changed from red (90% of trials) to green (10% of trials). Additionally, on each trial, a visual stimulus was presented in the background which participants were told was

irrelevant for their fixation task. Three blocks, termed "image-category localizer" blocks, included images as the irrelevant stimulus. In the other three blocks, termed "Gabor-orientation localizer" blocks, irrelevant stimuli were Gabor patches with a gray square in their center. Block-specific details are described below. Blocks were presented in an alternating order, with the first block type counterbalanced across participants.

During three image-category localizer blocks, background stimuli were images of food (50%) and vehicles (50%), obtained from the MemCat image dataset (Goetschalckx & Wagemans, 2019). Images were presented in the center of the screen ~8.5 cm for 800 ms with a 200 ms inter-trial interval (ITI), 600 trials per block. The fixation dot turned green, requiring participant responses, on 10% of trials, half of which were presented during simultaneous presentation of food images and half during vehicle images. Images were always trial-unique within a block but could be presented up to two times throughout all three image-category localizer blocks. Image presentation order was randomized within blocks.

Gabor-orientation localizer blocks featured background stimuli composed of Gabor patches ~27.5 cm wide with a ~8.5 cm gray square in their center. Gabor patches were oriented horizontally on half of trials and vertically on the other half of trials, presented in a random order. Patches were presented for 800 ms with a 200 ms ITI, 600 trials per block. The fixation dot changed from red to green on 10% of trials, half of which included horizontally-oriented Gabor patches and half of which included vertically-oriented Gabor patches.

Participants were able to rest between blocks. For both image-category and Gabor-orientation localizer tasks, fixation color change detection was performed as an attention check task. Trials on which the dot turned green were excluded from EEG analyses.

*Continuous performance task*

Following the EEG localizer task, participants performed four blocks of a continuous performance task (CPT) designed to test fluctuations in sustained attention. This task was a modified version of a not-X CPT during which participants are instructed to respond with a button press to frequent-category stimuli and withhold their button to rare targets (Robertson et al., 1997). Stimuli in the current study always comprised an image (~8.5 cm) superimposed in the center of a Gabor patch (~27.5 cm). Stimuli were presented for 800 ms, followed by a 500 ms blink period during which participants were instructed to blink—to reduce the number of blink artifacts during trials—and lastly a 200 ms ITI. Each block included 600 trials total. During two blocks, termed "Image-relevant CPT" blocks, participants were instructed to make a judgement about the category of the image and Gabor patches were task-irrelevant. During the other two blocks, "Gabor-relevant CPT" blocks, participants made responses based on the orientation of the Gabor patches and images were task-irrelevant. Task details for each block are described below. All images presented during the CPT were trial-unique and were not presented in any other localizer or CPT block. Image- and Gabor-relevant blocks were performed in alternation, with the first block counterbalanced across participants and unique from the order of the localizer task.

During image-relevant CPT blocks, images were drawn from one of three image categories. Two were frequent categories—food images (45% of trials) and vehicle images (45%). The third category—animal images (10% of trials)—served as rare targets. Participants were instructed at the beginning of the block to respond by pressing the space bar when they saw images of food or vehicles, and to withhold their response when the image was an animal image. Responses to frequent-category images occurring before the onset of the subsequent trial were

considered correct. Task-irrelevant Gabor patches were oriented either horizontally (50% of trials) or vertically (50% of trials).

Gabor patches during Gabor-relevant CPT blocks were oriented in one of three directions. Frequently, patches were oriented vertically (45% of trials) or horizontally (45% of trials). On rare target trials (10%), patches were oriented diagonally with a 45-degree left tilt. Participants were instructed to respond by pressing the space bar when patches were oriented horizontally and vertically but withhold their response when the patch was oriented diagonally. Responses to horizontal and vertical Gabor patches occurring before subsequent stimulus onset were considered correct. Task-irrelevant images were either food images (50% of trials) or vehicle images (50% of trials).

*Two-alternative forced choice memory task*

Finally, participants completed a surprise two-alternative forced choice (2AFC) memory task for image stimuli presented during CPT blocks. On each trial, one old and one new image were presented to the left and right of fixation, presentation order counterbalanced across trials. Participants were instructed to report the identity of the old image and provided a confidence rating—definitely the image on the left, maybe the image on the left, maybe the image on the right, definitely the image on the right—using the z, x, n, and m keys, respectively. The new image was always a lure from the same image category. Task instructions remained on the screen for all trials and trials were self-paced such that participants could take as long to respond as needed. Memory was tested for 60 relevant food images, 60 relevant vehicle images, 60 relevant animal images, 60 irrelevant food images, and 60 irrelevant vehicle images, resulting in a total of 300 memory trials. Relevant image memory was always tested first, followed by irrelevant image memory. EEG data

from the memory task is not analyzed as participants were not required to use the chin rest during this task. One participant chose to leave prior to the memory task. Therefore, memory analyses include data from 24 participants.

*EEG data collection*

EEG data was collected using 30 active electrodes (actiCHamp Plus, Brain Products, Munich, Germany) within an electrically shielded Faraday cage. Electrodes were placed at the following sites: Fp1, Fp2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2. The ground electrode was placed at Fpz. Two electrodes were placed on the right and left mastoids. During data collection, electrodes were referenced to the right mastoid. Offline, data were re-referenced to the algebraic average of the left and right mastoids. Data were filtered online with a high cutoff of 140 Hz and digitized at a sampling rate of 500 Hz using BrainVision Recorder (Brain Products, Munich, Germany) running on a PC.

Electrooculogram (EOG) activity was recorded using passive electrodes. Horizontal EOG activity was recorded using electrodes placed approximately 1cm from the outer canthus of each eye. Vertical EOG activity was recorded using a pair of electrodes placed above and below the right eye. A ground electrode was placed on the left cheek.

*Eye tracking*

Gaze position and pupil size were collected using a desk-mounted infrared eye-tracking system (EyeLink 1000 Plus, SR Research, Ottawa, Ontario, Canada) sampled at 1000 Hz. A chin

rest helped participants maintain a stable head position. The eye tracker was calibrated at the start of each EEG localizer and CPT block.

*Preprocessing and artifact rejection*

EEG data were bandpass filtered between 0 and 80 Hz and epoched from 150 ms prior to trial onset to 800 ms after trial onset. Baseline correction was applied using the time period -150 to 0 ms prior to stimulus onset.

Artifact rejection was performed using an automated detection procedure on EEG and EOG data. Trials containing artifacts were excluded from EEG analyses but retained for behavioral analyses. Due to participant motion over long task blocks, gaze position and pupil data was lost for large portions of task blocks making it unsuitable for artifact rejection. However, gaze position and pupil data were retained for control analyses.

Trials containing saccades or blinks were rejected if any of the following criteria were met: 1) if the difference in the mean activity in the first and second halves of an 80 ms sliding window (advanced in 10 ms increments) exceeded 60 microvolts in EEG channels or 50 microvolts in EOG channels; 2) if the range of activity within a 200 ms window exceeded 100 microvolts in EEG channels or 300 microvolts in EOG channels, or 3) if the absolute value of channel activity exceeded a difference from baseline of 100 microvolts in EEG channels or 300 microvolts in EOG channels. Trials containing linear drifts were excluded if the slope of a linear fit across the epoch exceeded 75 microvolts and the coefficient of determination ($R^2$) exceeded 0.3. Finally, trials with dropout or in which any channel flatlined for at least 200 ms were excluded.

Known noisy electrodes, frontal EEG channels Fp1 and Fp2, were excluded from preprocessing and subsequent analyses for all participants. In the case that noise in a single

additional EEG channel resulted in excessive artifacts, that channel was excluded from preprocessing and all subsequent analyses.

*Participant exclusion*

Participants were excluded from analyses if they met any of the following criteria, established in the preregistration: 1) Average behavioral performance (sensitivity, *A'*) during CPT blocks fell below 2.5 SD of the group average, or 2) More than half of CPT trials were excluded due to artifact rejection. No participants were excluded for low CPT performance. Four participants were excluded for excessive artifact rejection. The final sample size included 25 participants (mean age=22.24 years, SD=2.93 years) The gender breakdown was as follows: 15 men, 9 women, 1 non-binary person.

*Behavioral methods*

Individual-level behavioral performance on the CPT and during the localizer was quantified using *A'*, a non-parametric measure of sensitivity. The formula for calculating *A'* is as follows, where *hits* represents the hit rate and *FAs* represents false alarm rate:

If hits > FAs: $A' = 0.5 + (((hits-FAs) * (1+hits-FAs))/(4*hits*(1-FAs)))$

If FAs > hits: $A' = 0.5 - (((FAs-hits) * (1+FAs-hits))/(4*FAs*(1-hits)))$

Sensitivity *(A')* was calculated for each block separately and averaged within participants across image-relevant and Gabor-relevant blocks to obtain a measure of performance per block type.

Frequent responding during the CPT provides a behavioral index of sustained attentional state throughout the duration of the task. Extensive previous work has linked momentary measures of response time speed and variance to sustained attention performance (Esterman et al., 2013; Robertson et al., 1997; Rosenberg et al., 2013). Specifically, periods of slow (deBettencourt et al., 2018; Wakeland-Hart et al., 2022) and consistent (Decker et al., 2023) responding are more likely to precede successful withholding of presses to rare targets. Further, these measures uniquely predict upcoming lapses (Corriveau, Chao et al., 2024; Corriveau et al., 2025) suggesting that they index separate components of attention-related behavior.

To obtain a continuous measure of response speed, time courses of responses to frequent, correct responses were linearly detrended within block to account for speeding over time and z-scored within block. To calculate a measure of response time variance, we computed the absolute value of detrended and z-scored response time courses to obtain a measure of deviation from the average response time. Finally, missing values were linearly interpolated from surrounding trials.

We aimed to test whether response speed and variance uniquely predicted upcoming sustained attentional lapses, i.e., presses to infrequent target stimuli, replicating previous work (Corriveau, Chao et al., 2024). We calculated the average response speed and variance in the three-trial window preceding an infrequent target. Using a Bayesian logistic mixed effects modeling approach, we fit models predicting accuracy on infrequent trials from pre-trial response speed, variance, their interaction, and a categorical predictor of CPT block type. Finally, a subject-level random intercept was included. Models were fit using the package brms in R using default priors (4 chains, 10,000 iterations each).

Individual-level memory performance was calculated as sensitivity (*A'*) across memory trials for relevant and irrelevant images separately. Images were considered correctly-remembered

if participants reported a confident, accurate judgement. Following our preregistration, we also report memory accuracy for both confident and less confident memory judgements. When calculating memory performance for confident judgements, hit rates were calculated as the number of old images confidently remembered divided by the total number of confident reports (hits and false alarms). False alarm rate was calculated as the number of confident false alarms divided by the total number of confident reports. Due to the two-alternative forced choice nature of the task, hit rates and false alarm rates always summed to 1.

We sought to examine what features influenced successful memory judgements in the current study. Primarily, we were interested in whether measures related to selective attention (task-relevance), salience (stimulus frequency), and sustained attention (pre-trial RT speed and variance) predicted memory when controlling for known predictors of memory such as stimulus memorability (Bainbridge, 2019), the likelihood of the paired memory e lure to be incorrectly reported as remembered (false-alarmability; Utochkin et al., 2025), and trial number. Memorability scores for all images in the MemCat dataset were provided by the authors (Goetschalckx & Wagemans, 2019). For the current analyses we used memorability without correction for false alarms. False-alarmability was calculated for memory lure images as the number of false alarms divided by the total number of presentations in the original dataset.

To test the influence of these features on memory we fit Bayesian logistic mixed effects models predicting memory accuracy from effects of task-relevance (i.e., whether an image was presented in an image-relevant or a Gabor-relevant CPT block), category frequency (i.e., whether or an image was a rare animal image), memory trial number, stimulus memorability, memory lure false-alarmability, and pre-trial speed, variance, and their interaction. We additionally included a subject-level random intercept term.

*EEG Classifiers*

The primary aim of this study was to test the impacts of attention on the representational strength of stimulus information as measured by EEG decoding performance. We hypothesized that both selective attention—i.e., whether a stimulus was task-relevant—and sustained attentional state—i.e., whether a stimulus was presented during a period of better sustained attention—would result in stronger EEG decoding performance as the result of more actively-mainted stimulus information in mind.

We trained within-subject EEG linear discriminant analysis (LDA) classifiers to decode image category (food vs. vehicles) or Gabor orientation (horizontal vs. vertical) on all EEG localizer block trials. Classifiers were trained on the raw EEG activity across channels at every time point. Training on these blocks ensured that data was completely independent from testing CPT blocks and removed any confound of responding on classifier training. Trained models were then tested on CPT block trials.

We calculated the average performance of classifiers as a function of selective attention by splitting trials into whether the classifier was predicting the category of task-relevant or task-irrelevant stimuli. To test the impacts of sustained attentional state, trials were split by whether they appeared in the highest quartile (in-the-zone state) or lowest quartile (out-of-the-zone state) of sustained attention using a quartile split on the response speed and variance time courses. Lastly, to test whether pre-trial stimulus representations predicted upcoming attentional performance, we examined classifier performance for the three trials immediately preceding an attentional lapse—i.e., an incorrect press to an infrequent target—versus those preceding a correctly-withheld response.

Primary analyses tested whether the time course of EEG decoding performance meaningfully differed as a function of attention when time-locking to the onset of the visual stimulus. To do so, we tested classifier performance at the same time point in the trial at which the classifier was trained. Performance for all EEG decoders was quantified using the area under the curve. This analysis allows us to test whether the time course of stimulus decoding across the trial differs as a function of attention.

We also predicted that meaningful differences in decoding accuracy may be present relative to when participants made a response to that stimulus. In other words, the time course at which information is present in neural signals relative to when they act on that information may differ under different selective and sustained attention contexts. Because no responses were made during the EEG localizer task during which classifiers were trained, we identified the time period during which EEG classifier performance was most stable across time and averaged the raw EEG activity during training trials within this time window. To do so, we performed a 3-fold cross-validated temporal generalization analysis in which EEG classifiers were trained and tested across all points in the trial. In this analysis, reliable off-diagonal classification indicates stable category information present in EEG signals over time.

We retrained classifiers on averaged signals and tested classifier performance across the entire CPT trial duration. Finally, we time-locked the decoding time course to the button press shifting time courses such that they were centered on the time a response was made. The shifted time course included 800 milliseconds prior to the response through 800 ms after the response. Missing values after shifting were filled with 0.5, representing chance-level decoding.

For comparison, we also performed k-fold cross-validated classification for both images and Gabor patches during the localizer task blocks. The EEG decoder was trained on two-thirds of

localizer trials and performance was tested on the left-out third of trials. This process was repeated until each third of trials served as the held-out set. Performance was averaged across the three folds to return the average decoding performance for each subject.

*Event related potentials*

To examine whether activity in response to the presentation of frequent stimuli differed under different sustained attentional states, we calculated the average activity in trials presented under good sustained attention (the three-trial window preceding a correctly-withheld response to an infrequent target) and poor sustained attention (the window preceding a commission error or lapse in response to an infrequent target) in each channel across the scalp. We then calculated the difference between these average activity time courses for each subject.

*Time-frequency analysis*

We performed a time-frequency decomposition analysis to examine differences in oscillatory power under different sustained attentional states. Specifically, we examined trials in the window preceding an infrequent target as a function of performance on that target trial. Using a set of posterior electrodes, P3/P4/Pz/O1/O2/Oz, we convolved EEG time series with a Morlet wave for frequencies between 2 and 30 Hz in 1 Hz steps. The number of cycles increased based on frequency using the calculation $n_{cycles}$=frequency/3, resulting in wavelets ranging from approximately 1 cycle at 3 Hz to 10 cycles at 30 Hz. This analysis was implemented using the *tfr_array_morlet* function for MNE in python. For each subject, we averaged the resulting power matrix across trials as a function of sustained attention and calculated the difference in the time-

frequency matrices. The resulting matrix demonstrates power differences in frequency bands between good and poor sustained attentional states.

*Decoding from eye data*

To examine whether any differences in EEG decoder performance may be driven by artifacts due to eye gaze or pupil size, we also trained LDA classifiers using the data collected by the infrared eye tracker (gaze position, pupil size) and raw activity from vertical and horizontal EOGs. Before training, missing values in eye data were replaced with the average feature value across all available EEG localizer trials (training dataset) or CPT trials (testing dataset). Eye classifiers were trained and tested identically to EEG classifiers to determine whether decoding time courses show similar modulation by attention. We compared predictions from classifiers trained on EEG and eye data by correlating trial-wise predictions at each time point.

*Predicting memory from EEG decoding*

Previous work suggests that memory for infrequent targets is predicted by the attentional state in which they are encountered, as measured behaviorally (Corriveau, Chao et al., 2024; Corriveau et al., 2025; Decker et al., 2023; deBettencourt et al., 2018; Wakeland-Hart et al., 2022) and neurally (Adam et al., 2015; Aly & Turk-Browne, 2016). Here, we examined whether EEG decoding performance, a hypothesized index of attentional state, predicts subsequent memory for infrequent target images. We tested this possibility by fitting Bayesian logistic mixed effects models predicting infrequent trial memory success from EEG classification confidence—that is, the probability that a food image was correctly classified as food, or a vehicle image as a vehicle—

in the three trials preceding its presentation during the CPT. Decoder confidence was averaged across the entire trial duration for each image. We additionally included variables expected to be related to image memory performance as controls: image memorability, paired lure false-alarmability, and memory trial number, as well as a random intercept of subject. Image memorability values were obtained from the shared MemCat image database (memorability without false alarm correction). We calculated paired lure false-alarmability using data shared in the MemCat image database. Specifically, false-alarmability was calculated as the number of false alarms made to a given image divided by the total number of times that image was tested.

*Differences from preregistration*

For transparency, we note changes from preregistered analyses here. First, Bayesian statistics are used in lieu of parametric significance testing when evaluating differences in decoding time courses throughout the manuscript.

We initially preregistered indexing sustained attentional state using a median split on both response speed and variance time courses. The decision to instead use upper and lower quartiles of both response speed and variance was made to improve power by isolating the most extreme sustained attentional states. However, results and conclusions do not differ between these analysis choices so we do not believe this decision impacts the interpretation of the current study.

To better explore differences in neural representational strength under different sustained attentional states, we analyzed decoding performance in the three-trial window prior to infrequent trials as an exploratory analysis. While rare, infrequent trials provide a more objective measure of sustained attentional state than response time indices. We believe that capturing objective lapses

in sustained attention enabled the differences observed in decoding accuracy in the preceding three-trial window.

Finally, one preregistered exploratory analysis proposed testing whether oscillatory power predicted successful memory for stimuli. However, we decided that a more relevant analysis would test differences in oscillatory power under better and worse sustained attention conditions, as the primary aim of this study was to examine neural signatures of attention. This analysis is reported in the main text.

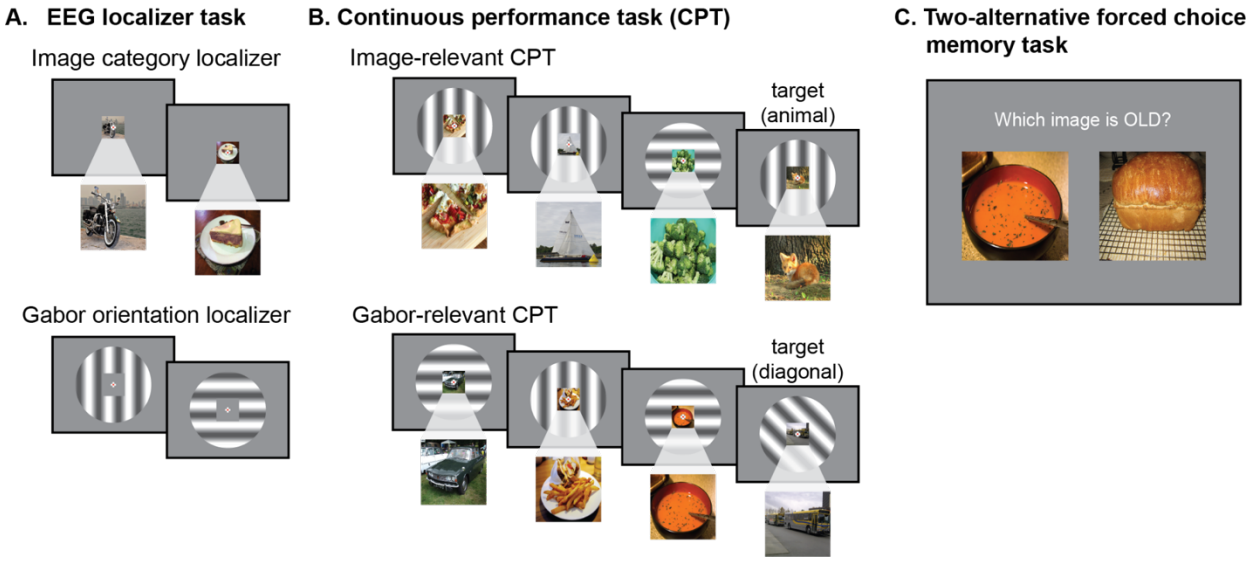



*Figure 28. Experimental procedure*. (A) Participants completed three tasks as part of the experimental procedure. The EEG localizer task included six alternating blocks of images (3 blocks) and Gabor patches (3 blocks). Participants detected a color change at fixation (red to green) which occurred on 10% of trials. Stimuli were presented for 800 ms with a 200ms inter-trial interval (ITI). Blocks included 600 trials each. (B) The continuous performance task featured stimuli composed of images superimposed on Gabor patches. Participants were assigned to make a judgement of image category (image-relevant runs) or Gabor orientation (Gabor-relevant runs),

pressing to frequent-category stimuli (food and vehicle images; horizontal and vertical Gabor patches) and withholding responses to rare targets (animal images; diagonal Gabors). Stimuli were presented for 800 ms with a 500 ms blink period and a 200 ms ITI. Participants completed four CPT blocks–two image-relevant blocks and two Gabor-relevant blocks in alternating order. (C) Finally, participants completed a surprise two-alternative forced choice memory task for images presented during CPT blocks. Participants were tasked with identifying the old image and provided a confidence rating (definitely or maybe) with a key mapping (z=definitely left; x=maybe left; n=maybe right; m=definitely right). This key mapping remained on the screen for all trials but is not depicted due to limited space. Participants completed 300 self-paced memory trials (180 relevant images, 120 irrelevant images).

**Results**

Participants (N=25, mean age=22.24 years, SD=2.93 years) performed a continuous performance task (CPT) which measures behavioral lapses in sustained attention while EEG recorded electrical activity across the scalp (**Figure 28**). In this modified version of the task, stimuli were images superimposed in the center of Gabor patches. To manipulate the focus of selective attention, participants were instructed at the start of each task run to make a frequency judgement about either the category of images or the orientation of Gabor patches, with frequent trials equally drawn from one of two categories (images of food or vehicles, horizontal or vertical Gabor patches) and infrequent trials drawn from a third rare category (images of animals, diagonally-oriented Gabor patches). Participants were tasked with responding to frequent-category stimuli (90% of trials) while withholding responses to infrequent-category stimuli (10% of trials). Sustained attention performance was measured discretely using accuracy on infrequent trials,

where correct omissions indicate good sustained attention and errors of commission indicate an attentional lapse. We also tracked continuous fluctuations in sustained attentional state using responses on correct, frequent trials by calculating time courses of response speed and variance which have been shown to foreshadow sustained attentional lapses (Corriveau et al., 2025; Corriveau, Chao et al., 2024; deBettencourt et al., 2018; Decker et al., 2023, Esterman et al., 2013; Rosenberg et al., 2013; Wakeland-Hart et al., 2022).

The aim of the current study was to test whether the classification of frequent-category stimuli (food vs. vehicle images; vertical vs. horizontal Gabor patches) from EEG signals differs as a function of attention, which would indicate differences in the amount of stimulus information available in brain signals under different attentional conditions. The task designed allowed us to compare unique contributions of both selective and sustained attention to differences in EEG classification performance by evaluating performance in task runs where stimuli were relevant vs. when they were irrelevant (selective attention), as well as when stimuli were presented during periods of better (in-the-zone) vs. worse (out-of-the-zone) sustained attentional states. To remove confounding factors of motor responses to EEG signals, within-subject linear-discriminant EEG classifiers were trained on separate task runs (EEG localizer task, see Methods) during which participants made a response only on rare, attention-check trials which were not included in classifier training. This study was preregistered at https://osf.io/h9r4e/overview. We report whether analyses were preregistered or exploratory as appropriate throughout the manuscript.

*Prolonged maintenance of task-relevant information in EEG signals*

Task performance (sensitivity, *A'*) was high in both the EEG localizer task used for classifier training (image runs: mean *A'*=.993, SD=$7.34*10^{-3}$, Gabor runs: mean *A'*=.993,

SD=8.27*10$^{-3}$) and the CPT used for classifier testing (image-relevant runs: mean *A'*=.941, SD=.027; Gabor-relevant runs: mean *A'*=.921, SD=.036), suggesting that participants generally maintained attention well. Participants also performed consistently during both the EEG localizer task ($r$=.795, p<.001) and the CPT ($r$=.728, p<.001), suggesting reliable individual differences in sustained attention performance. Because both image and Gabor patch stimuli were presented on all trials but varied by task-relevance, we were able to test whether the availability of stimulus information in EEG signals as measured with classification performance varied with the focus of selective attention. Our first registered analysis evaluated the time course of EEG classification performance throughout the trial duration to determine whether and when task relevance impacted how information was reflected in EEG activity. We hypothesized that neural representations of task-relevant stimuli would be more reliably reflected in EEG signals, enabling more successful decoding.

In line with our hypothesis, classification of task-relevant images (mean AUC=.554, SD=.022) from EEG signals was more accurate than classification of task-irrelevant images (mean AUC=.530, SD=.017) on average throughout the trial (t(24)=5.80, p<.001, $BF_{10}$>1000; **Figure 29A**). Examining the time course of image decoding, differences emerged between 300ms through 600ms after stimulus onset, with task-irrelevant image decoding dropping much more rapidly during this time, suggesting that image category information was not as well maintained in EEG activity. The initial rise in decoding performance did not differ by task-relevance, indicating consistent timing of image category classification regardless of task-relevance.

Performance for decoding the orientation of Gabor patches was low on average for relevant (mean AUC=.512, SD=8.98*10$^{-3}$) and irrelevant Gabor patches (mean AUC=.506, SD=8.80*10$^{-3}$). However, we observed moderate evidence for an effect of selective attention such that relevant

orientation decoding was higher than irrelevant decoding on average ($t(24)=2.83$, $p=9.31*10^{-3}$, $BF_{10}=5.08$; **Figure 29B**). Examining the time course of classification across the trial, task-relevant orientations remained higher late in the trial suggesting that information about Gabor orientations were maintained more strongly throughout the trial's duration when Gabor patches were the relevant stimulus. In sum, the attentional selection afforded to task-relevant stimuli leads to prolonged maintenance of stimulus information in EEG signals as indicated by stronger classification performance.

To determine whether the attentional selection afforded task-relevant stimuli increases the ability to decode stimulus information from EEG signals or whether task-irrelevance hinders this ability, we compare the decoding performance observed during CPT blocks to cross-validated classifier performance during EEG localizer blocks during which the classifier is trained. We observed that decoding performance for task-irrelevant stimuli was lower than cross validated performance during the localizer task, suggesting that task-irrelevant stimulus information is more strongly suppressed during the CPT (**Appendix 4 Figure 48**).

Finally, we examined whether decoding performance was driven by eye movements rather than signals derived from neural sources. We retrained classifiers to decode image categories and Gabor orientations using eye gaze and pupil data (see Methods). The time course of decoding from eye-based data is distinct from that of EEG-trained classifiers (**Appendix 4 Figure 49**). Further, predictions from EEG- and eye-based classifiers are not differentially correlated between relevant and irrelevant predictions, suggesting that eye-related features do not drive the differences observed as a function of task relevance (**Appendix 4 Figure 49**).

*Individual differences in decoding accuracy tradeoffs*

Previous work examining memory for relevant and irrelevant stimuli suggests no evidence of a tradeoff in processing as a function of task-relevance (Corriveau, Chao et al., 2024; Corriveau et al., 2025). However, by examining memory alone, past work has been unable to test evidence (or lack) of a processing tradeoff at the moment of encoding. Therefore, in an exploratory analysis, we leveraged trial-level classifier predictions to test whether classification performance for relevant and irrelevant stimuli within the same trial are reliably related across individuals. We correlated average decoding performance for relevant and irrelevant stimuli across all trials within a subject. A significant positive correlation on average across participants would suggest that processing co-fluctuates simultaneously for relevant and irrelevant stimuli. Alternatively, a negative correlation would suggest a tradeoff such that increased processing of relevant stimuli is accompanied by decreased processing for irrelevant stimuli presented on the same trial, and vice versa.

On average, we did not find evidence for either a simultaneous boost nor a tradeoff in image-relevant (mean $r_{relevant,\ irrelevant}$=-.056, SD=.332; One-sample t-test: t(24)=-.822, p=.420, $BF_{10}$=.286) and Gabor-relevant runs (mean $r_{relevant,\ irrelevant}$=-.045, SD=.316; One-sample t-test: t(24)=-.694, p=.495, $BF_{10}$=.262). This result adds to growing evidence that, on average, processing for relevant information does not increase at the expense of irrelevant processing, or vice versa. Results did indicate a wide range across participants of the extent to which information processing was either positively or negatively correlated. Therefore, we also tested whether this processing tradeoff, or lack thereof, was consistent within an individual. We correlated the level of tradeoff between relevant and irrelevant information for image-relevant and Gabor-relevant runs across participants. Interestingly, we observed strong consistency across participants ($r$=.964, p<.001)

suggesting that the level of processing tradeoffs is reliable within an individual. In other words, individuals who show simultaneous boosts in processing for relevant and irrelevant information do so regardless of the relevant stimulus category.

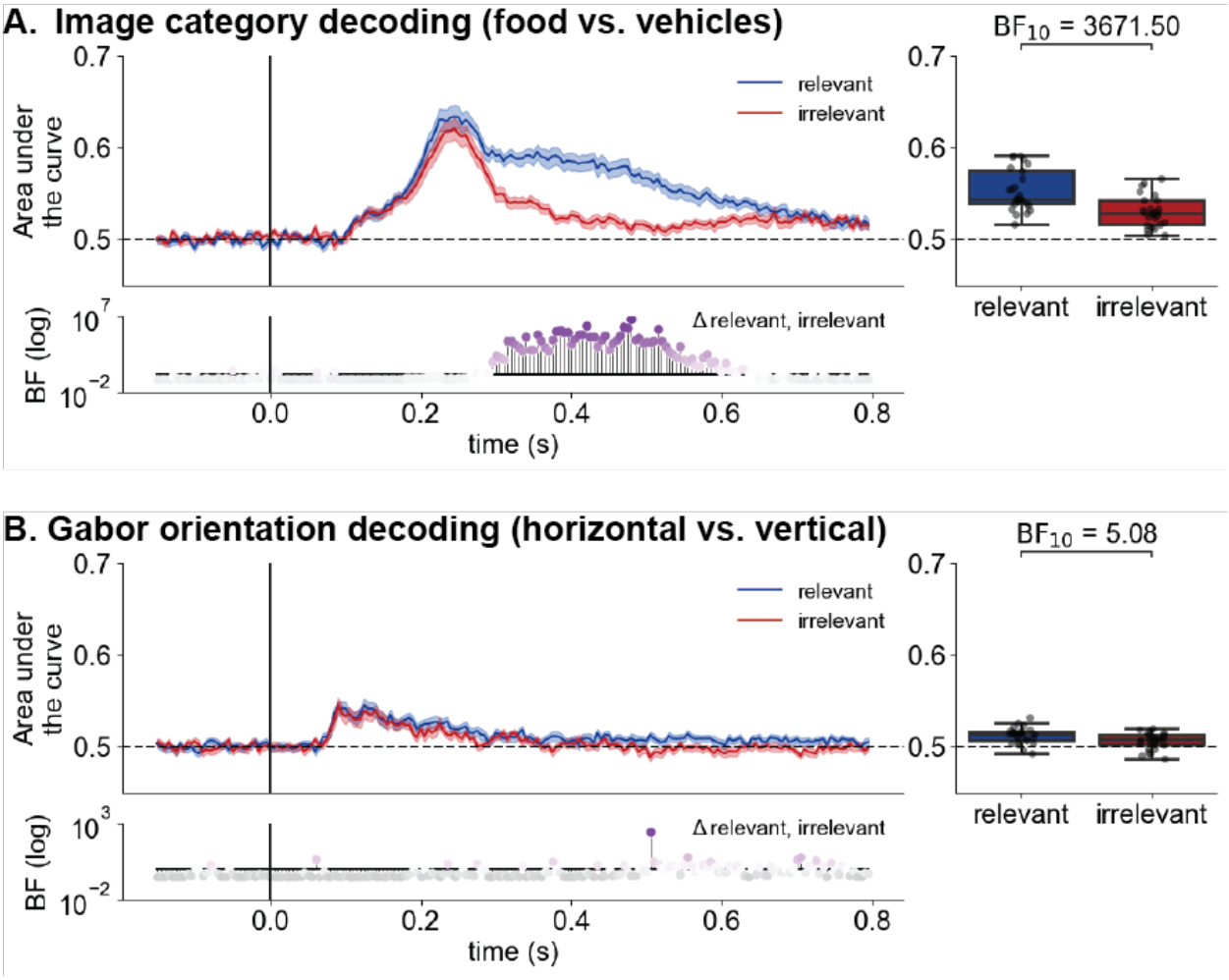



*Figure 29. Selective attention strengthens decoding of task-relevant stimuli.* The classification of (A) image categories and (B) Gabor orientations during a continuous performance task is higher when stimuli are task-relevant. Bayes factors (BFs) reflect evidence of differences between task-relevant and task-irrelevant time courses across participants. BFs greater than one are shown in

color and indicate evidence of a difference. Box-and-whisker plots reflect subject-wise mean area under the curve across the trial (0-800 ms after onset).

*Sustained attention lapses are preceded by degraded EEG representations*

Having established an effect of selective attention, i.e., stronger decoding for task-relevant information, we next tested whether fluctuations in sustained attentional state impacted trial-wise decoding performance. Infrequent targets provide a momentary probe into sustained attentional state such that correct performance (successfully-withheld responses to rare stimuli) indicates a high sustained attentional state whereas a commission error indicates lapsing attention. The following exploratory analysis tested whether CPT target performance was foreshadowed by differences in representational strength in EEG signals in the three-trial window preceding the presentation of a rare target. This trailing window approach has previously been shown to predict upcoming sustained attention lapses both behaviorally (Corriveau et al., 2025; Corriveau, Chao et al., 2024; deBettencourt et al., 2018; deBettencourt et al., 2019; Decker et al., 2023; Shelat & Giesbrecht, 2026; Wakeland-Hart et al., 2022) and neurally (deBettencourt et al., 2015). Due to poor decoding performance of Gabor orientations as predicted in our preregistration, we focus on the time course of image category decoding and report Gabor orientation decoding results in the supplement only (**Appendix 4 Figure 50**). Here, we tested how decoding of both relevant and irrelevant image category information differed based on whether participants subsequently lapsed during the presentation of an infrequent target.

Relevant images presented prior to a sustained attentional lapse (mean AUC=.529, SD=.047) were less accurately decoded relative to those shown prior to a correct omission indicating a good sustained attentional state (mean AUC=.561, SD=.028) suggesting further

influence of sustained attention fluctuations on EEG representations ($t(24)=4.04$, $p<.001$, $BF_{10}=152.2$; **Figure 30A**). This difference began ~300ms after onset and persisted for the remainder of the trial. These results indicate that trial-wise fluctuations in sustained attention affect the availability of stimulus information in EEG signals such that lapses are foreshadowed by degraded representations of relevant content.

We did not observe differences in the average decoding of irrelevant image categories as a function of sustained attentional state (mean $AUC_{correct}=.532$, $SD_{correct}=.033$; mean $AUC_{lapse}=.543$, $SD_{lapse}=.040$; $t(24)=-1.15$, $p=.263$, $BF_{10}=0.38$; **Figure 30B**). Across the time course, small differences emerge in irrelevant image decoding performance, with pre-lapse decoding briefly and inconsistently outperforming performance under better sustained attentional states. However, the lack of reliable differences in performance suggests that fluctuating sustained attentional states have minimal effect on representations of irrelevant stimuli under the conditions tested here.

Are differences in decoding performance preceding better vs. worse sustained attention performance driven by differential evoked EEG activity under different states? We next examined whether the onset of images prior to correctly-withheld responses elicited different levels of raw EEG activity than those prior to sustained attentional lapses by examining the differences in evoked response potentials (ERPs) across channels. We observed little to no evidence of ERP differences across the scalp preceding correct vs. lapsing sustained attention performance (**Appendix 4 Figure 51**). This suggests that the differences in decoding performance observed under different sustained attentional conditions were not due to overall differences in evoked amplitudes, but rather differences in the ability to detect multivariate signals under better vs. worse sustained attentional states.

Finally, as an exploratory analysis, we tested whether oscillatory power, in particular posterior alpha power, differed between trials preceding correctly-withheld responses (good sustained attention) and attentional lapses by examining the differences in the time-frequency decomposition between states. Posterior alpha power (8-12 Hz), as well as theta power (4-8Hz) was greater on trials preceding sustained attentional lapses around the time of stimulus onset (**Appendix 4 Figure 52**) as well as prior to trial offset. This suggests that sustained attentional states are characterized by differences in oscillatory power in low frequency bands and these differences are most pronounced at key moments in a trial (e.g., stimulus onset and offset).

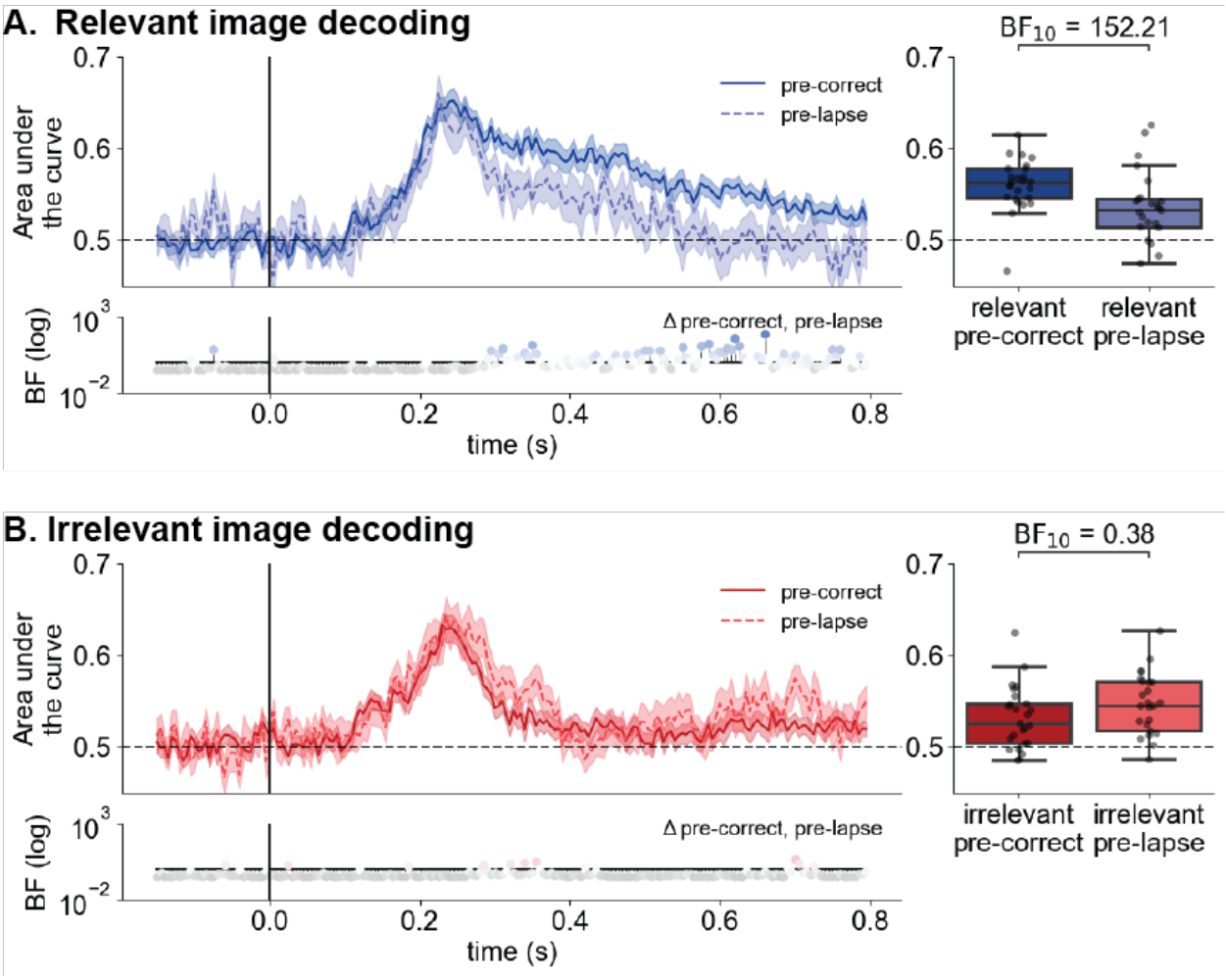



*Figure 30. Weak neural representations of relevant information prior to sustained attentional lapses.* Decoding performance was poorer for (A) relevant stimuli preceding sustained attentional lapses suggesting that neural representations are weaker under worse sustained attentional states. (B) Irrelevant image decoding did not differ by sustained attentional state on average. Box-and-whisker plots reflect subject-wise mean area under the curve across the trial (0-800 ms after onset).

*Response time measures predict attention lapses but not EEG decoding differences*

Because participants made responses on the vast majority of CPT trials, we are able to monitor continuous changes in sustained attentional states through measures related to response

time. Specifically, previous work has robustly linked lapses in sustained attention to periods of speeded responses (deBettencourt et al., 2018; Wakeland-Hart et al., 2022) as well as highly variable, erratic responding (Esterman et al., 2013; Rosenberg et al., 2013). In the current study, Bayesian mixed effects modeling with effects of response speed, variance, their interaction, and block type (image-relevant or Gabor-relevant) found evidence that behavioral responding predicted performance on infrequent sustained attentional probes during the CPT. Specifically, correctly-withheld responses were preceded by slower (mean effect=0.515, 89%CI [0.448, 0.584) and less variable pressing (mean effect=-0.181, 89%CI [-0.239, -0.122]). We also found evidence of an interaction between response speed and variance (mean effect=-0.053, 89%CI [-0.077, 0.031]) such that fast but less-variable responses resulted in better performance than fast but variable responses. Finally, participants made fewer attentional lapses in image-relevant CPT blocks (mean effect=0.473, 89%CI [0.373, 0.575]). Results replicate previous work suggesting that behavioral indices of speed and variability are uniquely related to trial-level sustained attentional performance.

We predicted that differences in decoding performance would emerge as a function of sustained attentional state as measured with response speed and variance. While our preregistered analyses described performing median splits based on response speed and variance, we opted to analyze the most upper and lower quartiles of trials (i.e., those occurring under the most extreme sustained attentional states as determined by response measures) to provide the most power for detecting a difference. Even when analyzing trials in the best and worst sustained attentional states as indicated by response speed (Relevant images: mean $AUC_{in\text{-}zone}$=.553, $SD_{in\text{-}zone}$=.027; mean $AUC_{out\text{-}of\text{-}zone}$=.556, $SD_{out\text{-}of\text{-}zone}$=.029; t(24)=-.493, p=.626, $BF_{10}$=.236; Irrelevant images: mean $AUC_{in\text{-}zone}$=.531, $SD_{in\text{-}zone}$=.021; mean $AUC_{out\text{-}of\text{-}zone}$=.528, $SD_{out\text{-}of\text{-}zone}$=.023; t(24)=.550, p=.587,

$BF_{10}$=.242) and variance (Relevant images: mean $AUC_{in\text{-}zone}$=.555, $SD_{in\text{-}zone}$=.028; mean $AUC_{out\text{-}of\text{-}zone}$=.550, $SD_{out\text{-}of\text{-}zone}$=.027; t(24)=.871, p=.392, $BF_{10}$=.297; Irrelevant images: mean $AUC_{in\text{-}zone}$=.532, $SD_{in\text{-}zone}$=.024; mean $AUC_{out\text{-}of\text{-}zone}$=.529, $SD_{out\text{-}of\text{-}zone}$=.024; t(24)=.549, p=.588, $BF_{10}$=.242), we did not observe differences in trial-averaged image category decoding for either relevant or irrelevant images (**Figure 31**). Timepoint-by-timepoint differences were brief and inconsistent, suggesting that the current study design lacked the ability to detect differences in stimulus representations as a function of continuous attentional state changes. We similarly observed no differences in Gabor orientation decoding as a function of response-indexed sustained attention across the trial (**Appendix 4 Figure 53**).

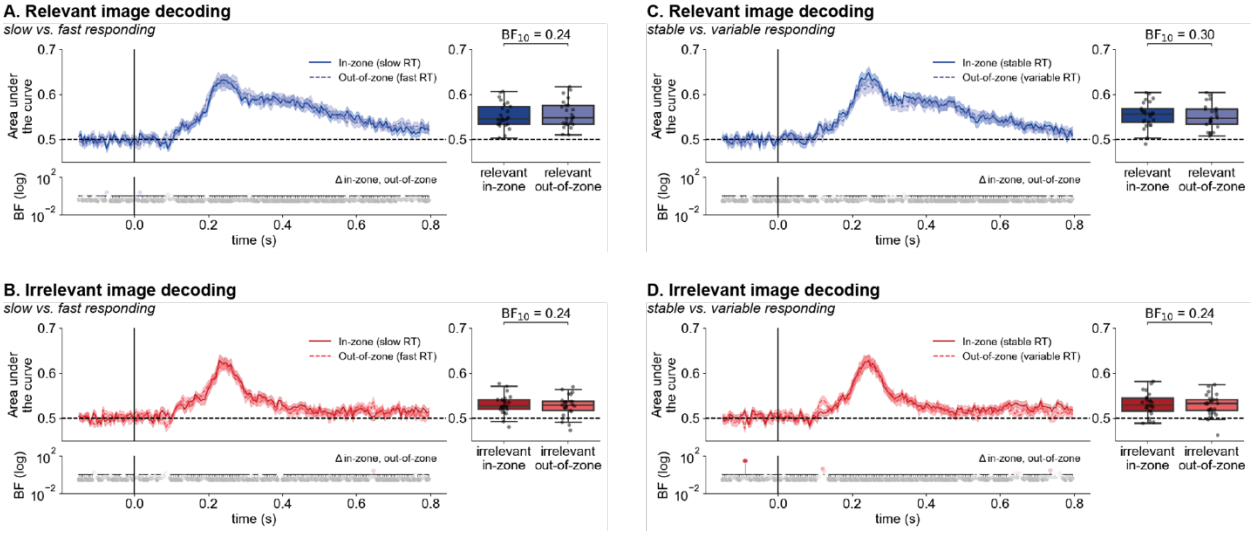



*Figure 31. Response time measures of sustained attention do not impact decoding.* Decoding performance of relevant and irrelevant images did not reliably differ as a function of sustained attentional states as determined by response time (RT) speed (A-B) nor variance (C-D). Box-and-whisker plots reflect subject-wise mean area under the curve across the trial (0-800 ms after onset).

*Confident EEG decoding predicts successful memory*

Attention, in concert with many stimulus- and task-related factors, impacts what we remember (Corriveau et al., 2026; Sherman & Turk-Browne, 2024). In the current study, we were interested in whether attention-related measures affected which stimuli were subsequently remembered, in addition to features such as stimulus memorability and false-alarmability which are known predictors of memory (Bainbridge et al., 2019; Utochkin et al., 2025). To test memory in the current study, participants performed a two-alternative forced choice memory task for a subset of images presented during CPT blocks.

Our first exploratory analysis tested what confluence of features impact image memory. We constructed a combination model predicting confident image memory from measures of selective attention (task-relevance), stimulus frequency, and response time predictors of sustained attention. We included additional predictors expected to be related to image memory: memorability, paired lure false-alarmability, and trial number, as well as a subject-level random intercept. We report the posterior distribution mean of estimated effects, where distributions whose 89% credible interval do not include zero are considered evidence of an effect. We also repeated this analysis predicting memory for both confident and unconfident memory judgements, i.e., judgements in which participants correctly selected the old stimulus regardless of confidence report. Results from this model are reported in the supplement (**Appendix 4 Table 14**).

When predicting confident memory judgements for all stimuli tested, we observed strong evidence for effects of stimulus memorability (mean= 0.291, 89%CI [0.237, 0.345]), paired lure false-alarmability (mean=-0.141, 89%CI [-0.192, -0.091]), and memory trial number (mean=-0.402, 89%CI [-0.495, -0.309]), suggesting that those stimuli with higher memorability scores, those paired with lures with lower false-alarmability scores, and those presented earlier during the

memory task were likely to be remembered. In addition, we observed infrequent targets (animal images) were better remembered than frequent-category images (mean=1.34, 89%CI [1.212, 1.459]), in line with work showing better memory for rare stimuli (von Restorff, 1933; Wallace, 1965). Effects of task-relevance (mean=0.082, 89%CI [-0.121, 0.288]), RT speed (mean=0.064, 89%CI [-0.0044, 0.132]), and RT variance (mean=-0.028, 89%CI [-0.089, 0.033]) were in the expected direction, but the credible intervals of these posterior distributions intersected zero. We saw no evidence of the interaction between RT speed and variance in relation to confident memory judgements (mean=-0.012, 89%CI [-0.043, 0.016]). In sum, both stimulus-related and task-related features impacted successful image recognition.

Next, following our finding that decoding performance predicted upcoming lapses in the CPT, we were further curious as to whether stronger decoding also foreshadowed successful memory for infrequent target images. To test this question, we analyzed decoding performance, i.e., the classifier-predicted probability that a stimulus belonged to its true image category, in the three-trial window preceding the presentation of an infrequent stimulus. We fit a Bayesian mixed effects logistic regression model predicting confident target memory from effects of decoding performance averaged across the trial, as well as memorability, paired false-alarmability, and memory task trial number. We also included a random intercept of subject. This analysis was proposed as an exploratory analysis in our preregistration. Again, we report results from a model predicting both confident and non-confident memory in the supplement (**Appendix 4 Table 15**).

As expected, target memory was predicted by image-level memorability (mean=.045, 89%CI [.024, .066]), the likelihood of the paired lure to elicit false alarms (mean=-.059, 89%CI [-.080, -.038]), and trial number (mean=-.052, 89%CI [-.072, -.031]). Additionally, EEG decoding performance positively predicted memory (mean=.025, 89%CI [.0045, .046]) suggesting that the

availability of stimulus information in EEG signals may index an encoding-ready state. Taken together, successful memory ultimately depends on a number of unique factors, including one's internal state which is reflected in the reliability of stimulus representations in brain signals.

## Discussion

The current study provides novel insights into the impacts of both selective and sustained attention on neural representations of visual information as measured by EEG decoding performance. Using a modified sustained attention task, we dissociated the effects of moment-by-moment sustained attentional states from task-relevance which was manipulated at the block level. We found effects of both selective and sustained attention on the decoding time course of image category information indicating that these attentional subcomponents uniquely impact the representation of stimulus information in neural signals. We additionally examined features that predicted recognition memory for images, showing that both attention-, task-, and stimulus-related features affect mnemonic judgements. Results provide new insights into the multifaceted nature of attention and its relationship with memory.

Decoding performance of EEG classifiers was reliably modulated by task-relevance. During image-relevant blocks, in which participants made responses based on image category, EEG decoding of image category (food vs. vehicles) remained strong for an extended duration relative to blocks in which image category information was task-irrelevant despite similar visual displays in both tasks, suggesting prolonged maintenance of task-relevant image information. This difference emerged following an initial shared peak in decoding performance for both relevant and irrelevant images, indicating that representations of low-level category information may not differ with task-relevance. Additionally, despite low decoding performance overall during Gabor-

relevant blocks, orientation decoding was stronger when Gabor patches were the task-relevant stimuli rather than when they were incidentally-presented. Results align with previous work demonstrating that selective attention may update attentional templates during visual search such that relevant information can be more easily decoded from EEG signals (Battistoni et al., 2020; Kaiser et al., 2016). Current results suggest that, even when spatial arrangement remains constant, task-relevant information is more-strongly represented in brain signals, enabling better decoding.

EEG classifiers were trained on data from runs involving passive viewing of stimuli. Despite this, decoding performance when stimuli were task-relevant did not exceed cross-validated performance within localizer (training) runs. From this result, we conclude that selectively attending stimulus information during the CPT did not strengthen category or orientation information in signals, but rather that task-irrelevance led to the suppression of this information. However, future work may seek to train a classifier on attended stimuli to test whether selective attention during training impacts generalization across tasks.

On top of task-relevance, we found that dynamic fluctuations in sustained attentional state influences the time course of decoding. Specifically, we observed worse decoding performance late in the trial for task-relevant images presented prior to sustained attentional lapses. This suggests that neural representations of task-relevant images in EEG signals are more degraded under poor sustained attentional states, in line with hypotheses of decoupling during periods of worse sustained attention (Schooler et al., 2011; Shelat & Giebrecht, 2025). To our knowledge, this is the first demonstration of weakened representational fidelity accompanying objective lapses in sustained attentional state.

Despite our observation that response speed and variance predicted sustained attention performance as measured with behavior (accuracy on infrequent trials), EEG classification results

did not support our prediction that decoding performance would differ by attentional zone, as indexed by response time measures. Stimuli presented during periods of slower and more stable responding, indicative of in-the-zone sustained attention (deBettencourt et al., 2018; Esterman et al., 2013), were not better-classified than those showing during periods of fast, variable responding. This lack of a difference likely reflects robust but noisy indexing of sustained attentional state with response time measures. Future work may seek to boost experimental power (e.g., number of trials) to uncover neural signatures of sustained attention and measured with response speed and variance.

We observed low decoding performance overall for Gabor orientation information, even when Gabor patches were task-relevant. While poor decoding was expected, as noted in our preregistration, we hypothesized that sustained attentional state would impact decoding performance. However, we did not observe reliable differences as a function of sustained attentional state for orientation decoding. Poor orientation decoding is likely due to the presentation of Gabor patches in the periphery of the visual field, for which weaker decoding performance has been previously noted (Kandemir & Olivers, 2024). Future work may seek to test how arrangement of the visual display interacts with selective and sustained attention effects on decoding performance.

Finally, we examined the impact of stimulus-, task-, and attention-related features on memory performance. As expected, previously-reported features like stimulus-specific memorability and lure false-alarmability predicted whether an image would be successfully recognized (Bainbridge et al., 2019; Utochkin et al., 2025). Interestingly, we found that memory for target images was predicted by classifier performance in the window preceding their presentation. This suggests that EEG classification success, which tracks sustained attention, may

reflect a processing state that enables successful mnemonic encoding. These results further support a key link between dynamic attentional and mnemonic states and emphasize the importance of considering attention in models of memory (Corriveau et al., 2026).

In sum, the current study tested the unique contributions of distinct attentional constructs—sustained and selective attention—on the time course of visual stimulus representations as measured by EEG decoding accuracy. We observed prolonged maintenance of task-relevant neural representations, as well as degraded representations preceding lapses in sustained attention. Finally, we observed that EEG classification performance predicted subsequent image memory, demonstrating that neural signals reliably index fluctuating attention and encoding-ready states. Taken together, results dissociate the impacts of attentional components on neural representations and highlight their role in ongoing informational processing.

**Chapter 6 Bibliography**


Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The Contribution of Attentional Lapses to Individual Differences in Visual Working Memory Capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616. https://doi.org/10.1162/jocn_a_00811

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

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

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2020). The time course of spatial attention during naturalistic visual search. *Cortex, In Honour of Dr Robert Rafal*, *122*, 225–234. https://doi.org/10.1016/j.cortex.2018.11.018

Bracci, S., Daniels, N., & Op de Beeck, H. (2017). Task Context Overrules Object- and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex*, *27*(1), 310–321. https://doi.org/10.1093/cercor/bhw419

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience*, *33*(39), 15454–15465. https://doi.org/10.1523/JNEUROSCI.2472-13.2013

Chidharom, M., Jones, H. M., Rosenberg, M. D., & Vogel, E. K. (2025). *Decoding Distraction From the Human Brain: A Unique Neural Signature Beyond Failures of Selective Attention and Control* (p. 2025.09.24.678372). bioRxiv. https://doi.org/10.1101/2025.09.24.678372

Corriveau, A., Chao, A. F., deBettencourt, M. T., & Rosenberg, M. D. (2025). Recognition memory fluctuates with sustained attention regardless of task relevance. *Psychonomic Bulletin & Review*, *32*(2), 714–728. https://doi.org/10.3758/s13423-024-02560-x

Corriveau, A., James Jr., A. R., deBettencourt, M. T., & Rosenberg, M. D. (2025). Sustained attentional state is a floodlight not a spotlight. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001769

Corriveau, A., Ke, J., & Rosenberg, M. D. (2026). Common Brain Network Dynamics Capture Attention Fluctuations in Tasks and Movies. *Journal of Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/JOCN.a.2622

Corriveau, A., Rosenberg, M. D., & deBettencourt, M. T. (2025). *Cognitive Neuroscience of Attention and Memory Dynamics* (N7tma_v1). PsyArXiv. https://doi.org/10.31234/osf.io/n7tma_v1

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

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., Keene, P. A., Awh, E., & Vogel, E. K. (2019). Real-time triggering reveals concurrent lapses of attention and working memory. *Nature Human Behaviour*, *3*(8), 808–816. https://doi.org/10.1038/s41562-019-0606-6

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

deBettencourt, M. T., Williams, S. D., Vogel, E. K., & Awh, E. (2021). Sustained Attention and Spatial Attention Distinctly Influence Long-term Memory Encoding. *Journal of Cognitive Neuroscience*, *33*(10), 2132–2148. https://doi.org/10.1162/jocn_a_01748

Decker, A. L., Duncan, K., & Finn, A. S. (2023). Fluctuations in Sustained Attention Explain Moment-to-Moment Shifts in Children's Memory Formation. *Psychological Science*, *34*(12), 1377–1389. https://doi.org/10.1177/09567976231206767

Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *Journal of Neuroscience*, *35*(36), 12383–12393. https://doi.org/10.1523/JNEUROSCI.1134-15.2015

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., 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

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

Goetschalckx, L., & Wagemans, J. (2019). MemCat: A new category-based image set quantified on memorability. *PeerJ*, *7*, e8169. https://doi.org/10.7717/peerj.8169

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences*, *111*(10), E962–E971. https://doi.org/10.1073/pnas.1312567111

Jackson, J., Rich, A. N., Williams, M. A., & Woolgar, A. (2017). Feature-selective Attention in Frontoparietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *Journal of Cognitive Neuroscience*, *29*(2), 310–321. https://doi.org/10.1162/jocn_a_01039

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience*, *31*(22), 8210–8219. https://doi.org/10.1523/JNEUROSCI.6153-09.2011

Jones, H. M., Yoo, K., Chun, M. M., & Rosenberg, M. D. (2024). Edge-Based General Linear Models Capture Moment-to-Moment Fluctuations in Attention. *Journal of Neuroscience*, *44*(14). https://doi.org/10.1523/JNEUROSCI.1543-23.2024

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *Journal of Neuroscience*, *36*(41), 10522–10528. https://doi.org/10.1523/JNEUROSCI.1385-16.2016

Kandemir, G., & Olivers, C. (2024). Comparing Neural Correlates of Memory Encoding and Maintenance for Foveal and Peripheral Stimuli. *Journal of Cognitive Neuroscience*, *36*(9), 1807–1826. https://doi.org/10.1162/jocn_a_02203

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

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. *Cerebral Cortex (New York, NY)*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

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., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., Ramani, R., Qiu, M., Constable, R. T., & Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, *117*(7), 3797–3807. https://doi.org/10.1073/pnas.1912226117

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics*, *75*(3), 426–439. https://doi.org/10.3758/s13414-012-0413-x

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

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319–326. https://doi.org/10.1016/j.tics.2011.05.006

Shelat, S., & Giesbrecht, B. (2025). Perceptual decoupling in the sustained attention to response task is likely: Comment on Bedi, Russell, & Helton (2024). *Experimental Brain Research*, *243*(4), 86. https://doi.org/10.1007/s00221-025-07032-9

Shelat, S., & Giesbrecht, B. (2026). Value-driven attentional capture prevents commission errors in real time. *Psychonomic Bulletin & Review*, *33*(5), 163. https://doi.org/10.3758/s13423-026-02930-7

Sherman, B. E., & Turk-Browne, N. B. (2024). Attention and Memory. In M. J. Kahana, A. D. Wagner, M. J. Kahana, & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.21

Song, H., Shim, W. M., & Rosenberg, M. D. (2023). Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *eLife*, *12*, e85487. https://doi.org/10.7554/eLife.85487

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

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. (2009). Selecting for Memory? The Influence of Selective Attention on the Mnemonic Binding of Contextual Information. *Journal of Neuroscience*, *29*(25), 8270–8279. https://doi.org/10.1523/JNEUROSCI.1043-09.2009

Utochkin, I., Chiang, N., & Bainbridge, W. A. (2026). *Recognition memory asymmetries predicted by individual item memorability* (6n74d_v3). PsyArXiv. https://osf.io/preprints/psyarxiv/6n74d_v3/

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

Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M., & Esterman, M. (2021). Brain state-based detection of attentional fluctuations and their modulation. *NeuroImage*, *236*, 118072. https://doi.org/10.1016/j.neuroimage.2021.118072

Zuberer, A., Kucyi, A., Yamashita, A., Wu, C. M., Walter, M., Valera, E. M., & Esterman, M. (2021). Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *NeuroImage*, *229*, 117610. https://doi.org/10.1016/j.neuroimage.2020.117610

## Conclusion: Cognitive neuroscience of attention and memory dynamics

Attention and memory support nearly every aspect of daily life, from cooking breakfast to commuting to work to relaxing with a book. It is necessary to remember, for example, the ingredients in a recipe, the quickest route, and the key characters. Attention is required to monitor the pot on the stove, avoid hazards on the road, and notice central events in the story. These mental processes, which shape how individuals experience and navigate the world, have become two of the most studied topics in cognitive neuroscience. Despite the proliferation of attention and memory research, however, the importance of their dynamics has been underappreciated. That is, we do not attend, encode, and retrieve equally well over time. Rather, the degree of focus and likelihood of storing, maintaining, and retrieving a memory varies from one moment to the next. Therefore, it follows that not only does characterizing these moment-to-moment fluctuations reveal valuable insights about attention, memory, and their interactions, but that it is inadequate to understand attention and memory without also considering how they vary over time.

Accumulating evidence suggests that attention and memory processes fluctuate in tandem (**Figure 32**). These synchronous dynamics are evident in daily life. While reading a book, for example, attention lapses can interfere with the ability to remember details from a paragraph read just moments ago. On the flip side, recalling a significant argument between two characters may drive attentional engagement when they meet again. Indeed, attention covaries with working memory (deBettencourt et al., 2019), long-term memory encoding (deBettencourt et al., 2018), and long-term memory retrieval (Madore & Wagner, 2022). Characterizing the synchrony between attention and memory dynamics can thus reveal crucial links between these processes.

At the same time, there are also aspects of attention and memory that vary out of sync. While attentional fluctuations substantially impact which items are later recognized (deBettencourt

et al., 2018), they appear to be distinct from item memorability (Roberts & Pruin et al., 2025; Wakeland-Hart et al., 2022). Moreover, attention consists of multiple subcomponents, including sustained attention and selective attention, which may interact differentially with memory (Corriveau, Chao et al., 2025; Corriveau et al., 2024; deBettencourt et al., 2021; Mirjalili & Duarte 2025). Further, while attention performance progressively declines over long periods of watch (Mackworth et al., 1948), working memory capacity does not show comparable monotonic decrements over time (Adam et al., 2015; Hakim et al., 2020). As in these examples, identifying moments when attention and memory dynamics decouple can clarify the nature of their relationships.

The interactions between attention and memory are complex and multidirectional. Neither attention nor memory is a unitary construct and their constituent processes may interact in different ways at different times. Fluctuations in one attentional process may influence one type of memory but not another. Memory may guide attention in one situation but not another. Attention may impact memory encoding and retrieval which in turn may affect how attention is deployed in the future. One means of better untangling this web of interactions is identifying when attention and memory dynamics are shared and when they diverge. Moreover, monitoring dynamics in real time presents powerful opportunities for neuroadaptive designs, including closed-loop neurofeedback and real-time triggering, that can reveal causal associations by tracing and manipulating attention and memory. A fuller characterization of dynamics informs understanding of attention and memory individually as well as when and how they interact in the human mind and brain.

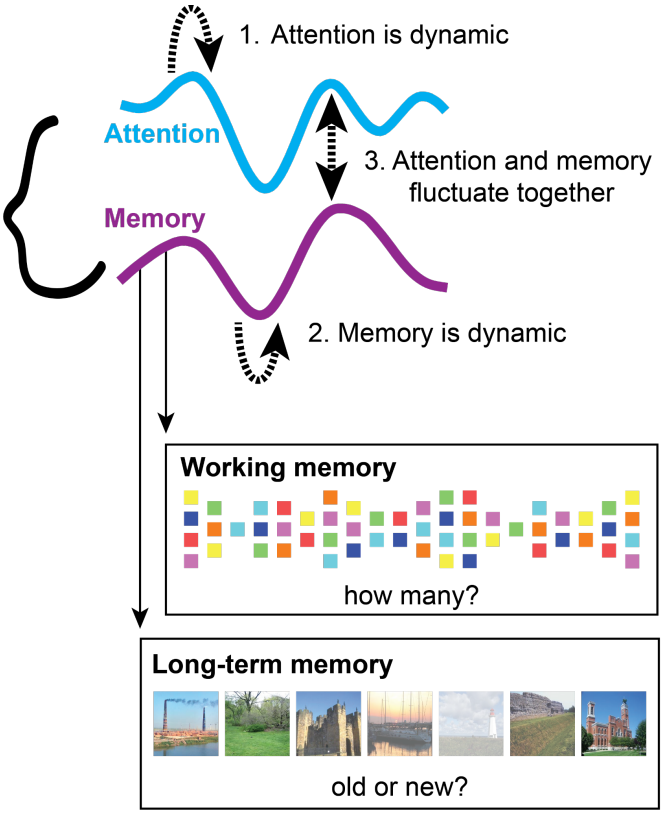



*Figure 32. Attention and memory each dynamically fluctuate over time.* Measuring these processes concurrently reveals that these fluctuations often occur in tandem, demonstrating profound links between sustained attention and memory, both working memory (how many items held in mind) and long-term memory (whether an item is recognized as old or new). Understanding when and which aspects of attention and memory share dynamics informs the mechanistic understanding of each of these processes as well as how they interact. A complete understanding of attention and memory processes requires studying how they evolve and co-evolve over time. Scene images are from the Scene Understanding Database (Xiao et al., 2010).

**Moment-to-moment fluctuations of attention**

Attention waxes and wanes from one moment to the next. Attentional states can be engaged one moment and disengaged the next. These moment-by-moment variations in attentional state may be driven by personal goals (e.g., top-down factors) or, for example, by something bright or loud in the environment capturing our focus (e.g., bottom-up factors). Even moments of better attention are multifaceted—the focus of attention can be directed at the external environment or one's own internal thoughts and experiences, guided by top-down goals or captured by salient information. The dynamic nature of attention may be most apparent in studies of sustained attention, which investigate how focus is maintained over time (Esterman & Rothlein, 2019; Fortenbaugh et al., 2017). In this section, sustained attention is leveraged as a case study for a broader understanding of attentional dynamics.

In sustained attention paradigms, participants are typically presented with a steady stream of stimuli. They are asked to detect rare targets (i.e., oddballs) that occasionally but infrequently occur during the stimulus stream. A key finding is that attention often lapses, as indexed by moments when participants fail to detect a rare stimulus. Continuous performance tasks (CPTs) provide a temporally rich readout of behavioral responses, including accuracy and response times (RTs). Notably, performance declines on sustained attention tasks over extended periods of time, an effect known as the vigilance decrement (Mackworth, 1948). Atop slow vigilance decrements, behavioral indices of attentional state (e.g., accuracy, RT speed, RT variability) also reveal how attention dynamically fluctuates over shorter timescales, ranging from seconds to tens-of-seconds (Castellanos et al., 2005; Rosenberg et al., 2025; Terashima et al., 2021). These behavioral indicators—including fast (Corriveau, Chao et al., 2025; deBettencourt et al., 2018) and erratic (Esterman et al., 2013; Rosenberg et al., 2013) responding—predict when attention failures are

most likely to occur. Although in many tasks fast RTs indicate engaged focus, in CPTs, speeding is a hallmark of automated, mindless responding. Erratic RTs likewise indicate suboptimal focus, potentially reflecting effortful performance more reliant on top-down control (Chidharom et al., 2025).

While response speed and variability forecast upcoming attention failures in CPTs, neurophysiological measures offer opportunities to examine brain activity that precedes and predicts lapses in a wider variety of tasks and contexts. In particular, cognitive neuroscience techniques with high temporal resolution, such as electroencephalography (EEG), can measure the oscillatory brain dynamics before lapses. Scalp EEG studies, for example, observe increases in alpha power (approximately 8-12 Hz) and decreases in theta power (approximately 4-7 Hz) before attentional lapses (Clayton et al., 2015; Mazaheri et al., 2009; O'Connell et al., 2009). This is consistent with observations that alpha power decreases prior to successful visual perception (Ergenoglu et al., 2004; Hanslmayer et al., 2007; van Dijk et al., 2008) and theta power increases to support cognitive control (Cavanagh & Frank, 2014). Studies measuring intracranial EEG can access activity in higher frequencies as well as in specific functional networks, revealing high frequency broadband (70-170 Hz) power increases in the default mode network (DMN) before attention lapses (Kucyi et al., 2020). Although functional magnetic resonance imaging (fMRI) has a much lower temporal resolution, comparisons of fMRI activity before lapses vs. successful responding have revealed reduced stimulus-specific activity and increased DMN, frontal (e.g., right inferior frontal gyrus, right middle frontal gyrus, and the anterior cingulate cortex), and parietal (e.g., left inferior parietal lobe, left precuneus, right supramarginal gyrus) activity seconds before as well as during lapses (Weissman et al., 2006). Finally, attention lapses are preceded by smaller pupils and linked to specific neuromodulatory states related to norepinephrine and locus

coeruleus function (Unsworth & Robison, 2016a). As such, behavioral and neurophysiological tools provide complementary insight into when a lapse is likely to occur.

Momentary lapses occur in the context of slower-fluctuating sustained attentional states (Rosenberg et al., 2025). These states—often discretized into "in-the-zone" states of stable, successful responding and "out-of-the-zone" states of erratic, error-prone performance (Esterman et al., 2013; 2014b; Rosenberg et al., 2013)—are characterized by univariate and multivariate differences in brain activity (deBettencourt et al., 2015; Esterman et al., 2013; Rosenberg et al., 2015). That is, although sustained attention task performance overall engages canonical attention regions, including the prefrontal and parietal cortices as well as subcortical structures (Corbetta & Schulman, 2002; Langer & Eickoff, 2013), the relative contribution of different brain regions and networks varies within in-the-zone and out-of-the-zone states.

Repeated evidence demonstrates that in-the-zone attentional states are characterized by increased DMN and decreased DAN activity (Esterman et al., 2013; Fortenbaugh et al., 2018; Kucyi et al., 2016; 2017) and co-fluctuation (a time point-by-time point measure of two regions' covarying activity; Jones et al., 2024). These attentional states also map onto DMN- and DAN-dominated brain states (patterns of activation and coactivation that cluster together in a low-dimensional neural representational space), respectively (Song et al., 2023; Yamashita et al., 2021). Although counterintuitive at first glance given the popular but simplistic conception of the DMN as a "mind wandering" or "off-task" network (Buckner et al., 2008; Christoff et al., 2009), increases in DMN activity are thought to reflect more automatic, practiced processing during in-the-zone states. In contrast, increased DAN activity in *worse* attentional states may reflect increased reliance on attentional control during these moments. First, during better attentional performance, these networks are more anticorrelated (Kucyi et al., 2020; Seeburger et al., 2024).

Second, interactions between brain regions and networks beyond the DMN and DAN reflect evolving attentional states (Song & Rosenberg 2021). Data-driven analytic approaches have revealed, for example, large-scale and distributed sets of functional connections whose dynamics predict attentional state changes during laboratory-based tasks (Jones et al., 2024; Kardan et al., 2022; Rosenberg et al., 2020; Zuberer et al., 2021) and naturalistic scenarios such as watching movies (Song et al., 2021). These networks include connections across the cortex, subcortex, and cerebellum and are not limited to functional connectivity within and between canonical networks such as the DMN and DAN alone (Rosenberg et al., 2016a). Finally, information processing differs between in-the-zone and out-of-the-zone states. The fidelity of neural representations increases under better attentional states (Jackson et al., 2017; Rothlein et al., 2018). Heightened sustained attention is likewise accompanied by greater evidence of distractor processing, suggesting that, in these periods, successful performance is possible without actively filtering task-irrelevant information (Esterman et al., 2014). Taken together, this evidence suggests that shifting attentional states are reflected in the activity and functional connectivity of widespread brain networks.

The behavioral and neuropsychological work discussed so far reveals correlates of attentional lapses and states. Causal approaches, however, offer a more direct way to explore the role of neural activity in sustained attention dynamics. For example, increasing attention via neuropharmacological manipulations also results in increased strength of and activity in networks correlated with better attention (Lyu, Corriveau, et al., in prep; Manza et al., 2025; Rosenberg et al., 2016b). The opposite is also true: decreasing attention with anesthetic agents results in decreased strength of attention-related networks (Chamberlain & Rosenberg, 2022; Rosenberg et al., 2020). In addition, methodological innovations enable researchers to monitor in real time the activity within specific brain networks (Sitaram et al., 2017; Stoeckel et al., 2014; Sulzer et al.,

2013). The time-varying dynamics of sustained attention can be fed back to participants in real time via closed-loop neurofeedback, thereby directly modulating brain activity and enhancing sustained attention performance (deBettencourt et al., 2015; Mennen et al., 2021). This real-time fMRI approach offers a promising avenue to directly test and modulate distributed patterns of functional connectivity that support sustained attention (Scheinost et al., 2020; Yamashita et al., 2017). By directly altering brain function, causal methods move beyond observation to actively test how neural representations support sustained attention behavior.

In sum, the dynamic nature of attention can be powerfully assayed through the lens of sustained attention. These studies reveal how attention varies over different time scales within individuals. Dynamics can be measured via behavior, including lapses, and tracked with cognitive neuroscience tools. These tools reveal that successfully sustaining attention is reflected in unique behavioral and neural signatures which can be monitored and leveraged causally to better understand the consequences of attentional changes over time.

**Moment-to-moment fluctuations of memory**

Much like sustained attention, memory performance fluctuates over time. For example, when meeting a new group of people, sometimes names are easily remembered while others---shared only seconds later---may be immediately forgotten. Such memory variability may be due to differences in the information itself (such as how salient or memorable it is; Bainbridge, 2019) as well as how it is encoded, maintained, or retrieved. Another non-exclusive possibility is that memory fluctuations are driven by shifting attentional, affective, and cognitive states that influence the ability to remember. Here, moment-to-moment variability in working memory and long-term memory is examined before considering how these constructs interact with fluctuating attention.

*Working memory dynamics*

From trial to trial, working memory performance changes. This is evident in variability across trials in both the number of items that can be reported from working memory (i.e., memory capacity; Adam et al., 2015; deBettencourt et al., 2019; Kozlova et al., PsyArXiv) as well as the fidelity of a single representation (i.e., memory precision; Zhang & Luck, 2008). At the same time, working memory tasks give rise to canonical brain signatures, including EEG delay signatures (Luria et al., 2016; Vogel & Machizawa, 2004) and distributed fMRI brain activity in stimulus-specific and control regions (Christophel et al., 2017; Curtis & D'Esposito 2003; D'Esposito & Postle 2015). By examining how these brain signatures covary with capacity and precision, cognitive neuroscience sheds light on the mechanisms that give rise to working memory dynamics. For example, working memory trials on which people remember fewer items are associated with reduced measures of preparatory activity (Murray et al., 2011) and delay activity in EEG (Adam et al., 2018) and fMRI (Pessoa et al., 2002) as well as smaller pupil size (Robison & Unsworth, 2019). More successful trials are also reflected in better multivariate representations (Bettencourt & Xu, 2016; Rahmati et al., 2018; Sprague et al., 2016; Wan et al., 2024). Causally implicating these neural substrates and oscillatory signatures are studies that demonstrate how transcranial magnetic stimulation (TMS) modulates working memory performance (Hamidi et al., 2008 & 2009; Riddle et al., 2020) which find, for example, that working memory accuracy increases relative to TMS-induced decreases in alpha power (Hamidi et al., 2009). In summary, working memory is inconsistent over time and investigating these inconsistencies can reveal important information about the neural underpinnings of working memory.

Although some momentary memory lapses—like forgetting someone's name as soon as they say it—can happen unexpectedly, there is evidence that often such failures are not isolated

events, but rather reflect ongoing dynamics shaped by both pretrial states and preceding content. Prior to stimulus presentation, neural activity, including posterior alpha (Myers et al., 2014) and frontal theta (Adam et al., 2015; 2018), predict upcoming working memory performance. That is, EEG provides access into general control states that may predetermine success. In addition, specific content from prior trials can influence working memory on an upcoming trial (Kiyonaga et al., 2017). Demonstrating the lingering effects of preceding trials on subsequent neural representations, evidence for prior stimulus representations has been decoded from neuroimaging and electrophysiological data, including EEG (Bae and Luck, 2019; Zhang & Lewis-Peacock, 2024), fMRI (St. John-Saaltink et al., 2016; Sheehan & Serences, 2022), MEG (Hajonides et al., 2023) and spiking activity (Barbosa et al., 2020). That is, like attention lapses, working memory failures likely reflect slower-frequency dynamics.

While working memory performance dynamics vary trial to trial—typically on the order of seconds to tens of seconds—a growing body of work investigates cognitive and neural dynamics *within* trials as information is held in mind during a delay period. This work reveals time-varying modulation of cognitive and neural stimulus representations on the order of milliseconds to seconds (Adam et al., 2022; Buschman & Miller, 2022; Stokes, 2015). Additionally, working memory behavioral performance, eye gaze, pupil size, and neural activity can be modulated while information is actively being held in mind during a delay interval by retrospective cues (Griffin & Nobre, 2003; Souza & Oberauer, 2016; van Ede et al., 2019; Zokaei et al., 2019). Thus, dynamics of working memory occur at a wide range of temporal scales. A comprehensive model of working memory must account for, and understand what gives rise to, these dynamics.

*Long-term memory dynamics*

One of the hallmark characteristics of long-term memory is that it is imperfect; we do not remember everything we encounter. For instance, how many new acquaintances' names do you remember hours, days, and weeks later? Experimentally, the stimulus properties, cognitive states, and neural signatures of this subsequent memory effect have been examined by contrasting brain activity and behavior from when items were initially encountered based on whether they were later remembered or forgotten (Paller & Wagner, 2002). Some variance in what is remembered can be driven by features of the item itself, for example, its valence or memorability (Bainbridge, 2019). However, memory is also critically shaped by internal states present when information is encountered, processed, and retrieved.

Lingering mnemonic states that influence whether an item will be remembered can be detected even before a stimulus appears. For example, memory judgments for an upcoming item can be biased by retrieval states from the previous trial (Duncan et al., 2012; Patil & Duncan, 2018). EEG provides powerful and temporally resolved insight into these prestimulus states. Using multivariate analysis of EEG, it is possible to decode retrieval states and predict memory outcomes (Long & Kuhl, 2019). In addition, oscillatory activity during the prestimulus period is associated with memory performance. Theta power prior to an item predicts long-term memory encoding (Fell et al., 2011; Fellner et al., 2013; Guderian et al., 2009; Merkow et al., 2014) and retrieval (Addante et al., 2011). Additionally, lower prestimulus alpha power at encoding is associated with successful long-term memory (Weidemann & Kahana, 2021). Prestimulus activity can also be measured by evoked brain activity, including EEG waveforms (Otten et al., 2006) and fMRI stimulus-related activity (Turk-Browne et al., 2006). In sum, prestimulus neural activity reflects ongoing mnemonic states that shape the likelihood of remembering.

An important observation is that mnemonic states unfold over longer timescales than single trials. This has been observed in long-term memory paradigms that test the ability to retrieve memories without explicit cues. For example, one task paradigm involves presenting participants a list of words and asking them to recall out loud all the words they remember from the list (i.e. free recall paradigms). A key finding from these tasks is that words presented close together in a list are also recalled nearby (Howard & Kahana, 2002; Manning et al., 2014; Polyn & Kahana, 2008). This behavior is attributable to slow fluctuations in internal context, measurable by multivariate analysis of EEG (Manning et al., 2011) and fMRI (Chan et al., 2017; Polyn et al., 2005). Items encoded simultaneously can also become contextually bound together (Gardner-Medwin, 1976; Horner & Burgess, 2013) and correlated with fMRI activity within the hippocampus (Horner et al., 2015). These dynamics may serve a key adaptive purpose, segmenting information that is presented over time into events (DuBrow et al., 2017; Shin & DuBrow, 2021). Thus, state dynamics over longer timescales critically shape whether and how information is remembered.

The identification of dynamic memory states presents a powerful opportunity for causal manipulation and real-time intervention. Memory encoding can be deliberately triggered during optimal or suboptimal moments, based on ongoing brain states (Rudoler et al., 2024; Salari and Rose, 2016; Yoo et al., 2012). Brain stimulation offers another way to manipulate memory (Hebscher & Voss, 2020; Yeh & Rose, 2019; Zhao & Woodman, 2021), and can be particularly powerful when paired with real-time decoding (Ezzyat et al, 2018; Kragel et al., 2025). Real-time decoding also enables adaptive interventions, such as prioritizing poorly encoded information for restudying (Fukuda & Woodman, 2015). In addition, real-time fMRI and neurofeedback can bias memory reactivation and retrieval (deBettencourt et al., 2019; Koizumi et al., 2016; Peng et al.,

2024; Taschereau-Dumouchel et al., 2018). Identifying and intervening upon mnemonic state dynamics offers a window into how memory can be shaped over time.

## Considering attention and memory dynamics together

While there is consensus that both attention and memory exhibit intrinsic fluctuations, most often these fluctuations are examined in isolation. The following sections describe work that combines concurrent measurements of attention and working and long-term memory to examine whether, when, and how their dynamics interact. Investigating these processes in tandem reveals their complex and evolving relationships.

### *Dynamics reveal that attention and working memory are tightly coupled*

Attention and working memory are closely conceptually related as working memory involves selective attention to perceptual information and internal representations of items held in mind (Adam & deBettencourt, 2019; Awh et al., 2006; Chun, 2011; Cowan 1998; Oberauer, 2019). Despite this close connection, most conceptualizations of attention and working memory also acknowledge that these constructs are not synonymous. Here, we first consider the relationship between attention and working memory through shared variance across individuals and similar neural representations. Then, we discuss how interleaving attention and working memory tasks provides critical insight into when and how their dynamics are intertwined.

Both attention and working memory abilities vary enormously across the population. Leveraging this variation reveals that attention and working memory abilities are correlated across individuals: Individuals who are better at sustaining attention also have higher working memory capacity (Kane et al., 2007; Unsworth et al., 2010; Unsworth & Robison, 2020). Neural analyses

also demonstrate how sustained attention and working memory are related across individuals. Functional-connectivity-based models trained to predict individual differences in one domain generalize to predict individual differences in the other (Avery et al., 2020; Kardan et al., 2022; Yoo et al., 2022). That is, behavioral and neural analyses of individual differences reveal a close link between sustained attention and working memory.

Another way to examine shared neural mechanisms is by testing whether attention and working memory engage similar cognitive and neural processes. This can be observed indirectly by the impact on performance of simultaneous dual attention and working memory tasks. Engaging in a distracting attention task while maintaining items in working memory reduces memory performance (Souza & Oberauer, 2017). Analyses of neural data can also reveal commonalities between these processes. First, neuroimaging and electrophysiological recordings from humans and non-human primates suggest that items elicit similar neural responses whether they are perceptually attended or maintained in working memory (Awh & Jonides, 2001; Gazzaley & Nobre, 2012; Ikkai & Curtis, 2011; Panichello & Buschman, 2021). Likewise, regions implicated in one process show overlapping activity in another: the frontal eye fields and intraparietal sulcus, for example, support goal-directed attention and are activated during working memory maintenance (Pessoa et al., 2002). Second, changes in common neural features predict fluctuations in *both* attention and working memory. For example, increased frontal theta power and decreased posterior alpha power predict better attention and working memory performance (Adam et al., 2015; 2018; Myers et al., 2014; van Ede et al., 2017).

Thus, attention and working memory abilities are correlated across individuals and also implicate similar neural substrates. In addition, both attention and working memory change across multiple time scales. But are those dynamics themselves linked? This question has been

challenging to answer, as dynamics have traditionally been observed using different stimuli in distinct paradigms. The most compelling way to answer this is by concurrently and even simultaneously tracking attention and working memory fluctuations. If attention and working memory co-fluctuate, this suggests an intricate link between them.

A study by deBettencourt et al., (2019) investigated whether sustained attention and working memory exhibit simultaneous dynamics by interleaving sustained attention and working memory tasks. Moments of better sustained attention (slower, more careful responding) correlated with moments of better working memory (more items remembered). Furthermore, this work presented a key innovation to directly leverage ongoing sustained attention. Fluctuations of sustained attention fluctuations were monitored in real time to specifically probe working memory when attention was extremely high or low (**Figure 33**). Participants remembered more items when probes appeared in a high vs. low attentional state, revealing that working memory and sustained attention dynamics are themselves deeply intertwined.

A follow-up study replicated these behavioral links between attention and working memory and explored pupillary correlates (Keene et al., 2022). This work also developed real-time pupil triggering, such that working memory probes were triggered when pupil size was exceptionally large or small. These observations that sustained attention and working memory covary together in time provide strong evidence that these constructs rely on a common resource store that dictates ability from moment to moment. Furthermore, the technological development of real-time triggering designs make important suggestions about how to leverage these dynamics to enhance attention and working memory behavior.

Examining attention and working memory dynamics in tandem also reveals dissociations in these processes. While attentional dynamics correlate with the *number* of items in working

memory on a given trial (capacity), they are not related to the *precision* of working memory representations (deBettencourt et al., 2019). This finding is in line with previous work that capacity and precision may comprise separable features of working memory (Fukuda et al., 2010; Murray et al., 2011). Furthermore, while attention and memory share some neural mechanisms, EEG (Hakim et al., 2019) and fMRI (Sheremata et al., 2018) findings also show dissociation in neural activity underlying these constructs. Indeed, working memory likely comprises multiple subcomponents, including attentional control (Adam et al., 2015; Hakim et al., 2020). Although attention can be deployed to representations in working memory, items need not be the focus of attention to be remembered (Olivers et al., 2011; Lewis-Peacock et al., 2012; LaRocque et al., 2015). Finally, subjective self-reports of moment-to-moment fluctuations of attentional states may be associated with working memory performance fluctuations (Unsworth & Robison, 2016b), but they may reflect distinct attentional states (Chidharom et al., 2025). Thus, examining the shared dynamics of attention and working memory can shed light on the intricate and complex relationship between the subcomponents of these processes.

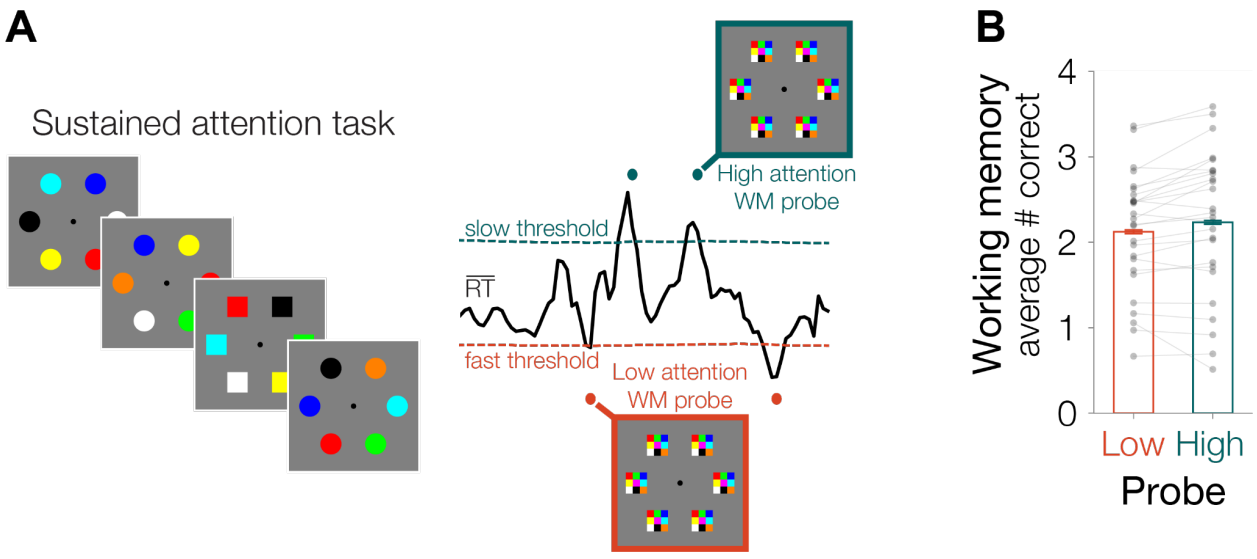



*Figure 33. Linking attention and working memory dynamics.* **(A)** A real-time triggering approach explores synchronous fluctuations in attentional state and working memory (WM) capacity. A sustained attention task was used to monitor fluctuations of attention in real time based on the speed of responding (fast or slow). This method enabled researchers to probe WM during moments of high and low sustained attention. Such real-time monitoring techniques are valuable tools for exploring cognitive dynamics and developing individualized and adaptive systems that respond to moment-to-moment changes in attentional state **(B)** WM performance was better during periods of better sustained attention, as participants recalled more colors correctly when attention was high ($p$<.01). These findings directly linked momentary lapses of sustained attention with the amount of information held in mind. Figure adapted from deBettencourt et al., 2019.

*Attention dynamics drive long-term memory*

Although it may seem that paying better attention to something will help it be remembered later on, little work has tested this relationship empirically. Rather, understanding how fluctuating attentional states drive memory has been a challenge for the field. Here we will examine the relationship between sustained attention and long-term memory through their shared dynamics.

Attention and long-term memory share common sources of variability at both behavioral and neural levels. Across individuals, higher sustained attention is correlated with better long-term memory performance (Corriveau, Chao et al., 2025; deBettencourt et al., 2021), though this relationship may be particularly strong among children and young adults (Decker et al., 2023b; Tran et al., 2025). Factor analyses suggest that sustained attention ability may, in fact, be more similar to long-term memory than to other attentional components like attentional control (Zhao et al., bioRxiv). Additionally, the recruitment of overlapping brain mechanisms for attention and long-term memory supports the close relationship between these functions (Aly & Turk-Browne 2017; Chun & Turk-Browne, 2007; Kuhl & Chun, 2014; Long et al., 2018). Retrieving locations from long-term memory activates similar multivariate representations as attending to spatial locations, evident in both EEG alpha topography (Sutterer et al., 2019) and fMRI activity in retinotopic regions (Vo et al., 2022). In addition, EEG activity patterns related to retrieval state are activated during a spatial attention task (Long 2023). Memory encoding recruits activity in the parietal cortex, specifically regions canonically associated with attention (Hutchinson et al., 2014; Turk-Browne et al., 2013; Uncapher et al., 2011; c.f. Hutchinson et al., 2009). On the other hand, attention is represented in the hippocampus, a region tightly linked with long-term memory (Aly & Turk-Browne 2016b; Cordova et al., 2019; Dudovic et al., 2011). Moreover, some of the shared

relationship between attention and memory may be mediated by reliance on shared neuromodulatory indices (Decker & Duncan, 2020; Tarder-Stoll et al., 2020).

Trial-level fluctuations of attention within individuals also predict variability in long-term memory performance (deBettencourt et al., 2021; Mirjalili & Duarte, 2025). Activity in frontoparietal attention network regions predicts successful long-term memory encoding (Turk-Browne et al., 2013), as do attention-related patterns of activity in the hippocampus (Aly & Turk-Browne, 2016a). Attentional state prior to retrieval, indexed by posterior alpha power and pupil size, predicts memory success (Madore et al., 2020). These findings all support the idea that long-term memory shares representations with attention more broadly, and particularly sustained attention.

More recently, research has directly examined how moment-to-moment fluctuations of sustained attention relate to long-term memory (**Figure 34**). Findings show that periods of better sustained attention lead to better memory (deBettencourt et al., 2018). This observation has since been replicated and extended by a growing body of research (Corriveau, Chao et al., 2025; Corriveau et al., 2024; Decker et al., 2023b; Wakeland-Hart et al., 2022). Findings robustly show that higher sustained attentional state, measured by more accurate, slower, and less variable responding, covaries with better recognition memory for stimuli that appear during those moments.

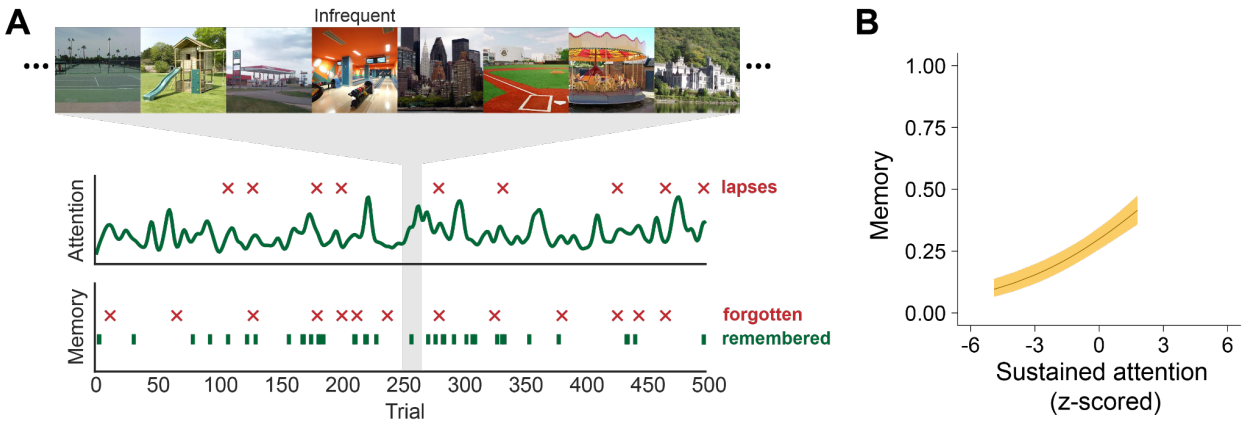



*Figure 34. Predicting long-term memory from moment-to-moment fluctuations in sustained attention.* **(A)** An experiment links sustained attention dynamics with long-term memory. Attentional state was tracked using trial-by-trial fluctuations in behavioral responses. Attention dynamics not only predicted lapses in attention in the moment but also critically related to which items were remembered or forgotten later on. **(B)** Recognition memory was significantly better on trials with higher attentional states ($p$<0.001), showing that attention dynamics strongly impact what is remembered. This suggests that even subtle fluctuations in attentional state play a profound role in memory formation. Figure adapted from Corriveau, Chao et al., 2025. Scene images are from the Scene Understanding Database (Xiao et al., 2010).

One study by deBettencourt et al., (2018) exploited attentional fluctuations to demonstrate the direct link between sustained attention and memory. Response speed was monitored in real time to determine moments when participants were most likely to be in an engaged state (showing slow RTs) or a disengaged attentional state (showing fast RTs). During these moments, probes were inserted requiring participants to change their prepotent response. Not only did the manipulation work—participants were more likely to lapse during worse attentional states—but memory was also worse for probe items presented in poor attentional states. This work was

expanded upon to show that attentional states are independent from memorability (Wakeland-Hart et al., 2022). Leveraging both attention and memorability in real time offers a powerful strategy to individually tailor which information is encoded when (Roberts & Pruin et al., 2025). These findings not only establish a direct link between attention and memory dynamics but also showcase the power of real-time, adaptive cognitive interfaces to dissect and unravel the complex relationship between attention and memory.

Attentional fluctuations at encoding also have residual effects on memory encoding of information outside the focus of attention. Work by Corriveau and Chao et al., (2025) presented pairs of images during the sustained attention task, one of which was task-irrelevant. Increases in attentional state not only predicted better memory for the task-relevant images but also for the irrelevant images. A conceptual replication of this work utilized multisensory audio-visual stimuli (Corriveau et al., 2024). Successful recognition of a task-relevant stimulus (e.g., a picture) predicted memory for its irrelevant pair (e.g., a sound), suggesting that sustained attentional state at encoding affected relevant and irrelevant stimuli similarly. Together, these findings suggest that better sustained attention does not sharpen filtering, but rather reflects a greater overall capacity of the attentional system.

Sustained attention paradigms, particularly their ability to reveal attention dynamics, are valuable for examining whether and how attention impacts memory and learning more broadly. For example, the attentional boost effect describes a phenomenon in which rare targets transiently improve memory for information encountered at the same time (Lin et al., 2010; Swallow & Jiang, 2010). The proposed mechanism for this effect suggests that detecting rare targets transiently and broadly boosts perceptual processing, facilitating encoding for the context that accompanies the target (Swallow & Jiang, 2013). However, while the detection of rare targets improves recognition

memory, it does not alter temporal memory (Wang & Egner, 2023). Also, sustained attention does not appear to impact temporal clustering of memory recall (Jayakumar et al., 2023). Finally, this approach also extends beyond memory, for example to examine the relationship with learning. There is evidence that attentional state impacts the ability to learn statistical regularities, with participants exhibiting faster responses for anticipated stimuli under high attentional states (Zhang & Rosenberg, 2024; although, see Decker et al., 2023a).

Just as attention affects what is remembered, memory similarly impacts attention. That is, knowledge and experience influence what, how, and when information is attended (Chen & Hutchinson, 2018; Nobre & Stokes, 2019). Work examining this relationship demonstrates that target detection is aided by contextual cueing from spatial arrangement or identity of distractors (Chun, 2000), memorized scenes (Summerfield et al., 2006), and learned temporal regularities (Cravo et al., 2017), suggesting that memories guide spatial attention. Additionally, distractors with learned value capture attention more than distractors with no value association (Anderson et al., 2011). Memory-guided attention is supported by the hippocampus (Aly & Turk-Browne, 2017) which represents (Gunseli & Aly, 2020) and differentiates (Favila & Aly, 2025) learned attentional states. Still underexplored, however, is the impact of memory on attentional dynamics. For example, do expectations formed from memories result in attentional state changes at specific moments in time? Related work demonstrates that temporal regularities in stimulus presentation decrease tonic pupil size, a proxy of attentional state (Shalev & Nobre, 2022), suggesting that learned temporal structure characteristically alters attentional dynamics. Additionally, sustained attention dynamics are sensitive to rewards, with higher accuracy and lower response variability following rewards (Trach et al., 2025). Future explorations should further test how memory of past experience shapes the trajectory of attentional state in time.

## Toward a dynamic framework for cognition

Attention and memory are two of the most studied constructs in cognitive psychology and neuroscience. However, this work often assumes both functions are stable and binary—i.e., information is either attended or unattended, remembered or forgotten—and disregards the fact that they fluctuate over time. Past work has been fruitful for characterizing attentional and mnemonic processes as a whole, and assuming stability may have been useful for identifying trademark properties of these functions. With this groundwork laid, future research should consider changes in attention and memory over time to more fully understand these inherently dynamic functions. In particular, methods incorporating cognitive dynamics may seek to answer a host of open questions.

### *How can dynamic interactions advance models of attention and memory?*

Incorporating the dynamic properties of both attention and memory into computational models offers new opportunities to better capture the complexity of cognition and behavior. For example, modeling attentional features improves computational models of memory performance (Adam et al., 2015; Hakim et al., 2020) as well as machine learning decoding (Mirjaili & Duarte, 2025). More broadly, fluctuating attentional states during encoding and retrieval may account for memory variability that has often been attributed to memory processes alone, suggesting a need to revise memory theories accordingly. Conversely, memory retrieval can shape future attentional deployment, implying that dynamic models of attention should also incorporate memory-based influences. In addition, attention and memory fluctuations are observed and change over an enormous range of timescales (Rosenberg et al., 2025), ranging from sub-second (i.e., theta) rhythms (Biba et al., PsyArXiv; Fiebelkorn & Kastner, 2019; Landau & Fries, 2012; Schwartz et

al., 2025) to the lifespan, as both attention and memory generally peak in early to middle adulthood (Alloway & Alloway, 2013; Fortenbaugh et al., 2015; Ronnlund et al., 2005). Modeling the bidirectional, time-varying interactions between attention and memory may provide a more unified framework for understanding human cognition. Joint measurement of attention and memory dynamics will be critical for uncovering the extent to which observed mnemonic and attentional phenomena emerge from shared underlying fluctuations, distinct mechanisms, or their interaction.

*How can understanding attention and memory dynamics inform technology and translation?*

A deeper understanding of attention and memory dynamics could have broad impacts, with applications across technology, education, and healthcare. Attention and memory have already inspired key innovations in artificial intelligence (Lindsay, 2020), and a fuller understanding and implementation of their dynamic properties may lead to more flexible and adaptive AI architectures. In education, attention and memory are essential for achieving learning goals, and tailoring instructional strategies to account for their natural fluctuations could significantly enhance educational effectiveness. In health care, attention and memory impairments are implicated in a range of psychiatric and neurological disorders, suggesting that tracking and modeling their fluctuations could support early detection and personalized interventions. Ultimately, insights into the dynamic nature of attention and memory could drive innovations across domains, shaping tools that are more responsive to the complexity of human cognition.

*How can brain-computer interfaces harness attention and memory dynamics in real time?*

Understanding the dynamic nature of attention and memory unlocks powerful opportunities for real-time prediction and intervention. Because cognitive states naturally fluctuate

over time, knowing the momentary state can offer strong predictive leverage for anticipating what comes next. When attention wanes, catastrophic lapses become more likely, working memory capacity shrinks, and long-term memory formation is impaired. Advanced signal processing and analytic frameworks allow researchers to track these fluctuations in real time using cognitive neuroscience and behavior tools (Shelat et al., 2024). Real-time monitoring of behavioral and neural signals enables researchers not just to observe, but to intervene in the moment. Approaches such as real-time triggering, neuroadaptive task designs, and closed-loop neurofeedback leverage continuous readouts of cognitive state to dynamically adjust stimuli, optimize performance, or directly modulate neural activity (**Figure 35)**. By moving beyond static snapshots and toward real-time dynamic monitoring, cognitive neuroscience can begin to both predict—and ultimately shape—the unfolding trajectory of mental processes.

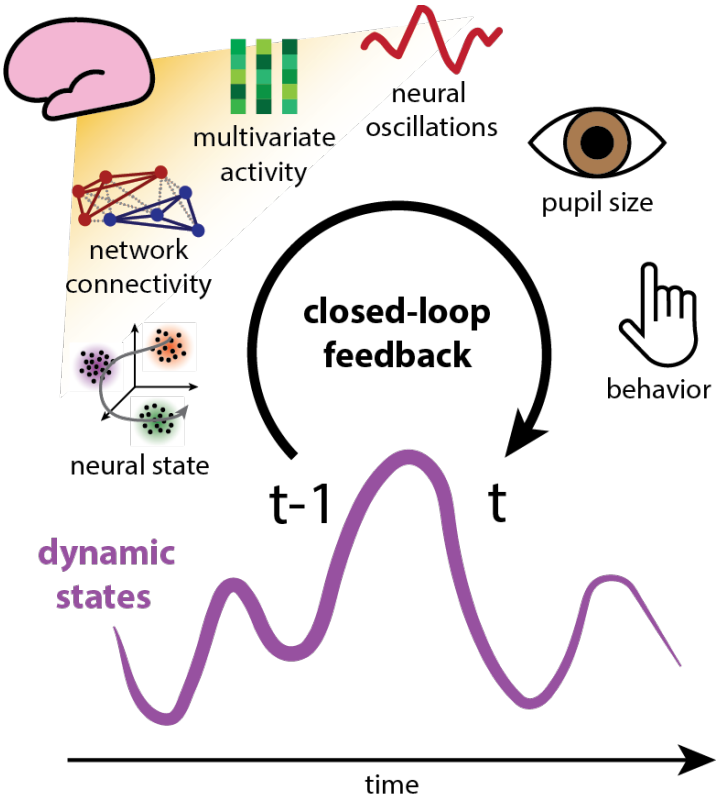



*Figure 35. Adaptive designs enable the tracking of dynamic states in real time.* Patterns of brain activity and behavior provide insight into ongoing processes which can be used to trigger timely interventions or guide real-time changes in the environment. Such closed-loop interfaces not only deepen understanding of the complex and intricate relationship between attention and memory but also offer a powerful framework for probing the dynamic nature of cognition itself.

## Conclusion

While we all have personal experience with what it feels like to attend and remember—as well as what it feels like when these abilities fail—the cognitive and neural dynamics underlying these phenomena are complex. Successful attention and memory may arise from some combination of external factors and drifting internal states whose interactions dictate ongoing cognition. Investigating fluctuations in attention and memory provides a critical lens into

understanding these processes in isolation, as well as their interaction. The advancement of cognitive neuroscience techniques enables the possibility of probing, leveraging, and perturbing brain states to better understand how slowly-evolving states affect cognition and behavior. Therefore, while the dynamics of attention and memory may shape daily experience, neuroscience may also hold the keys to shaping cognitive dynamics in turn.

**Conclusion Bibliography**


Adam, K. C. S., & deBettencourt, M. T. (2019). Fluctuations of Attention and Working Memory. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.70

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The Contribution of Attentional Lapses to Individual Differences in Visual Working Memory Capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616. https://doi.org/10.1162/jocn_a_00811

Adam, K. C. S., Rademaker, R. L., & Serences, J. T. (2022). Dynamics Are the Only Constant in Working Memory. *Journal of Cognitive Neuroscience*, *35*(1), 24–26. https://doi.org/10.1162/jocn_a_01941

Adam, K. C. S., Robison, M. K., & Vogel, E. K. (2018). Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance. *Journal of Cognitive Neuroscience*, *30*(9), 1229–1240. https://doi.org/10.1162/jocn_a_01233

Addante, R. J., Watrous, A. J., Yonelinas, A. P., Ekstrom, A. D., & Ranganath, C. (2011). Prestimulus theta activity predicts correct source memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(26), 10702–10707. https://doi.org/10.1073/pnas.1014528108

Alloway, T. P., & and Alloway, R. G. (2013). Working memory across the lifespan: A cross-sectional approach. *Journal of Cognitive Psychology*, *25*(1), 84–93. https://doi.org/10.1080/20445911.2012.748027

Aly, M., & Turk-Browne, N. B. (2016a). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), E420-429. https://doi.org/10.1073/pnas.1518931113

Aly, M., & Turk-Browne, N. B. (2016b). Attention Stabilizes Representations in the Human Hippocampus. *Cerebral Cortex (New York, N.Y.: 1991)*, *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. 369–403). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-50406-3_12

Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., Scheinost, D., Constable, T. R., & Chun, M. M. (2020). Distributed Patterns of Functional Connectivity Predict Working Memory Performance in Novel Healthy and Memory-impaired Individuals. *Journal of Cognitive Neuroscience*, *32*(2), 241–255. https://doi.org/10.1162/jocn_a_01487

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/s1364-6613(00)01593-x

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. https://doi.org/10.1016/j.neuroscience.2005.08.023

Bae, G.-Y., & Luck, S. J. (2019). Reactivation of Previous Experiences in a Working Memory Task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

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

Barbosa, J., Stein, H., Martinez, R. L., Galan-Gadea, A., Li, S., Dalmau, J., Adam, K. C. S., Valls-Solé, J., Constantinidis, C., & Compte, A. (2020). Interplay between persistent

activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nature Neuroscience*, *23*(8), 1016–1024. https://doi.org/10.1038/s41593-020-0644-4

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157. https://doi.org/10.1038/nn.4174

Biba, T., Decker, A., 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

Buschman, T. J., & Miller, E. K. (2022). Working Memory Is Complex and Dynamic, Like Your Thoughts. *Journal of Cognitive Neuroscience*, *35*(1), 17–23. https://doi.org/10.1162/jocn_a_01940

Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of Attention-Deficit/Hyperactivity Disorder-Related Intra-Individual Variability. *Biological Psychiatry*, *57*(11), 1416–1423. https://doi.org/10.1016/j.biopsych.2004.12.005

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Chamberlain, T. A., & Rosenberg, M. D. (2022). Propofol selectively modulates functional connectivity signatures of sustained attention during rest and narrative listening. *Cerebral Cortex (New York, N.Y.: 1991)*, *32*(23), 5362–5375. https://doi.org/10.1093/cercor/bhac020

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

Chidharom, M., Bonnefond, A., Vogel, E. K., & Rosenberg, M. D. (2025). Objective markers of sustained attention fluctuate independently of mind-wandering reports. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02640-6

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029

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

Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*, *19*(4), 188–195. https://doi.org/10.1016/j.tics.2015.02.004

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Córdova, N. I., Turk-Browne, N. B., & Aly, M. (2019). Focusing on what matters: Modulation of the human hippocampus by relational attention. *Hippocampus*, *29*(11), 1025–1037. https://doi.org/10.1002/hipo.23082

Corriveau, A., Chao, A. F., deBettencourt, M. T., & Rosenberg, M. D. (2025). Recognition memory fluctuates with sustained attention regardless of task relevance. *Psychonomic Bulletin & Review*, *32*(2), 714–728. https://doi.org/10.3758/s13423-024-02560-x

Corriveau, A., James, A., deBettencourt, M., & Rosenberg, M. (2024). *Sustained attentional state is a floodlight not a spotlight*. OSF. https://doi.org/10.31234/osf.io/k9cnm

Cowan, N. (1998). *Attention and Memory: An Integrated Framework*. Oxford University Press.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423. https://doi.org/10.1016/S1364-6613(03)00197-9

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., Keene, P. A., Awh, E., & Vogel, E. K. (2019). Real-time triggering reveals concurrent lapses of attention and working memory. *Nature Human Behaviour*, *3*(8), 808–816. https://doi.org/10.1038/s41562-019-0606-6

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

deBettencourt, M. T., Williams, S. D., Vogel, E. K., & Awh, E. (2021). Sustained Attention and Spatial Attention Distinctly Influence Long-term Memory Encoding. *Journal of Cognitive Neuroscience*, *33*(10), 2132–2148. https://doi.org/10.1162/jocn_a_01748

Decker, A., Dubois, M., Duncan, K., & Finn, A. S. (2023a). Pay attention and you might miss it: Greater learning during attentional lapses. *Psychonomic Bulletin & Review*, *30*(3), 1041–1052. 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., Duncan, K., & Finn, A. S. (2023b). Fluctuations in Sustained Attention Explain Moment-to-Moment Shifts in Children's Memory Formation. *Psychological Science*, *34*(12), 1377–1389. https://doi.org/10.1177/09567976231206767

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

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

Dudukovic, N. M., Preston, A. R., Archie, J. J., Glover, G. H., & Wagner, A. D. (2011). High-resolution fMRI reveals match enhancement and attentional modulation in the human medial temporal lobe. *Journal of Cognitive Neuroscience*, *23*(3), 670–682. https://doi.org/10.1162/jocn.2010.21509

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

Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. *Cognitive Brain Research*, *20*(3), 376–383. https://doi.org/10.1016/j.cogbrainres.2004.03.009

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 (New York, N.Y.: 1991)*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Esterman, M., Reagan, A., Liu, G., Turner, C., & DeGutis, J. (2014a). Reward reveals dissociable aspects of sustained attention. *Journal of Experimental Psychology. General*, *143*(6), 2287–2295. https://doi.org/10.1037/xge0000019

Esterman, M., Rosenberg, M. D., & Noonan, S. K. (2014b). Intrinsic fluctuations in sustained attention and distractor processing. *The Journal of Neuroscience: The Official Journal of the Society for 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

Ezzyat, Y., Wanda, P. A., Levy, D. F., Kadel, A., Aka, A., Pedisich, I., Sperling, M. R., Sharan, A. D., Lega, B. C., Burks, A., Gross, R. E., Inman, C. S., Jobst, B. C., Gorenstein, M. A., Davis, K. A., Worrell, G. A., Kucewicz, M. T., Stein, J. M., Gorniak, R., … Kahana, M. J. (2018). Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nature Communications*, *9*(1), 365. https://doi.org/10.1038/s41467-017-02753-0

Fell, J., Ludowig, E., Staresina, B. P., Wagner, T., Kranz, T., Elger, C. E., & Axmacher, N. (2011). Medial temporal theta/alpha power enhancement precedes successful memory encoding: Evidence based on intracranial EEG. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(14), 5392–5397. https://doi.org/10.1523/JNEUROSCI.3668-10.2011

Fellner, M.-C., Bäuml, K.-H. T., & Hanslmayr, S. (2013). Brain oscillatory subsequent memory effects differ in power and long-range synchronization between semantic and survival processing. *NeuroImage*, *79*, 361–370. https://doi.org/10.1016/j.neuroimage.2013.04.121

Fiebelkorn, I. C., & Kastner, S. (2019). A Rhythmic Theory of Attention. *Trends in Cognitive Sciences*, *23*(2), 87–101. https://doi.org/10.1016/j.tics.2018.11.009

Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, *1396*(1), 70–91. https://doi.org/10.1111/nyas.13318

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

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. https://doi.org/10.3758/17.5.673

Fukuda, K., & Woodman, G. F. (2015). Predicting and Improving Recognition Memory Using Multiple Electrophysiological Signals in Real Time. *Psychological Science*, *26*(7), 1026–1037. https://doi.org/10.1177/0956797615578122

Gardner-Medwin, A. R. (1976). The recall of events through the learning of associations between their parts. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *194*(1116), 375–402. https://doi.org/10.1098/rspb.1976.0084

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. https://doi.org/10.1016/j.tics.2011.11.014

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194. https://doi.org/10.1162/089892903322598139

Guderian, S., Schott, B. H., Richardson-Klavehn, A., & Düzel, E. (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5365–5370. https://doi.org/10.1073/pnas.0900289106

Hajonides, J. E., van Ede, F., Stokes, M. G., Nobre, A. C., & Myers, N. E. (2023). Multiple and Dissociable Effects of Sensory History on Working-Memory Performance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *43*(15), 2730–2740. https://doi.org/10.1523/JNEUROSCI.1200-22.2023

Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the Neural Focus of Attention Reveals Distinct Processes for Spatial Attention and Object-Based Storage in Visual Working Memory. *Psychological Science*, *30*(4), 526–540. https://doi.org/10.1177/0956797619830384

Hakim, N., deBettencourt, M. T., Awh, E., & Vogel, E. K. (2020). Attention fluctuations impact ongoing maintenance of information in working memory. *Psychonomic Bulletin & Review*, *27*(6), 1269–1278. https://doi.org/10.3758/s13423-020-01790-z

Hamidi, M., Slagter, H. A., Tononi, G., & Postle, B. R. (2009). Repetitive transcranial magnetic stimulation affects behavior by biasing endogenous cortical oscillations. *Frontiers in Integrative Neuroscience*, *3*. https://doi.org/10.3389/neuro.07.014.2009

Hamidi, M., Tononi, G., & Postle, B. R. (2008). Evaluating frontal and parietal contributions to spatial working memory with repetitive transcranial magnetic stimulation. *Brain Research*, *1230*, 202–210. https://doi.org/10.1016/j.brainres.2008.07.008

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bäuml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage*, *37*(4), 1465–1473. https://doi.org/10.1016/j.neuroimage.2007.07.011

Hebscher, M., & Voss, J. L. (2020). Testing network properties of episodic memory using non-invasive brain stimulation. *Current Opinion in Behavioral Sciences*, *32*, 35–42. https://doi.org/10.1016/j.cobeha.2020.01.012

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning & Memory*, *16*(6), 343–356. https://doi.org/10.1101/lm.919109

Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., & Wagner, A. D. (2014). Functional Heterogeneity in Posterior Parietal Cortex Across Attention and Episodic Memory Retrieval. *Cerebral Cortex*, *24*(1), 49–66. https://doi.org/10.1093/cercor/bhs278

Ikkai, A., & Curtis, C. E. (2011). Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia*, *49*(6), 1428–1434. https://doi.org/10.1016/j.neuropsychologia.2010.12.020

Jackson, J., Rich, A. N., Williams, M. A., & Woolgar, A. (2017). Feature-selective Attention in Frontoparietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *Journal of Cognitive Neuroscience*, *29*(2), 310–321. https://doi.org/10.1162/jocn_a_01039

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

Jones, H. M., Yoo, K., Chun, M. M., & Rosenberg, M. D. (2024). Edge-Based General Linear Models Capture Moment-to-Moment Fluctuations in Attention. *Journal of Neuroscience*, *44*(14). https://doi.org/10.1523/JNEUROSCI.1543-23.2024

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*(7), 614–621. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Kardan, O., Stier, A. J., Cardenas-Iniguez, C., Schertz, K. E., Pruin, J. C., Deng, Y., Chamberlain, T., Meredith, W. J., Zhang, X., Bowman, J. E., Lakhtakia, T., Tindel, L., Avery, E. W., Lin, Q., Yoo, K., Chun, M. M., Berman, M. G., & Rosenberg, M. D. (2022). Differences in the functional brain architecture of sustained attention and working memory in youth and adults. *PLOS Biology*, *20*(12), e3001938. https://doi.org/10.1371/journal.pbio.3001938

Keene, P. A., deBettencourt, M. T., Awh, E., & Vogel, E. K. (2022). Pupillometry signatures of sustained attention and working memory. *Attention, Perception & Psychophysics*, *84*(8), 2472–2482. https://doi.org/10.3758/s13414-022-02557-5

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences*, *21*(7), 493–497. https://doi.org/10.1016/j.tics.2017.04.011

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., & Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behaviour*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0006

Kozlova, O., Adam, K., & Fukuda, K. (2024). *Prospective and retrospective awareness of moment-to-moment fluctuations in visual working memory performance*. OSF. https://doi.org/10.31234/osf.io/nbftr

Kragel, J. E., Lurie, S. M., Issa, N. P., Haider, H. A., Wu, S., Tao, J. X., Warnke, P. C., Schuele, S., Rosenow, J. M., Zelano, C., Schatza, M., Disterhoft, J. F., Widge, A. S., & Voss, J. L. (2025). Closed-loop control of theta oscillations enhances human hippocampal network

connectivity. *Nature Communications*, *16*(1), 4061. https://doi.org/10.1038/s41467-025-59417-7

Kucyi, A., Daitch, A., Raccah, O., Zhao, B., Zhang, C., Esterman, M., Zeineh, M., Halpern, C. H., Zhang, K., Zhang, J., & Parvizi, J. (2020). Electrophysiological dynamics of antagonistic brain networks reflect attentional fluctuations. *Nature Communications*, *11*(1), 325. https://doi.org/10.1038/s41467-019-14166-2

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 of the United States of America*, *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 (New York, N.Y.: 1991)*, *27*(3), 1831–1840. https://doi.org/10.1093/cercor/bhw029

Kuhl, B. A., & Chun, M. M. (2014). Memory and attention. In *The Oxford handbook of attention* (pp. 806–836). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199675111.013.034

Landau, A. N., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900. https://doi.org/10.1037/a0030694

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

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Lin, J. Y., Pype, A. D., Murray, S. O., & Boynton, G. M. (2010). Enhanced Memory for Scenes Presented at Behaviorally Relevant Points in Time. *PLOS Biology*, *8*(3), e1000337. https://doi.org/10.1371/journal.pbio.1000337

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, *14*, 29. https://doi.org/10.3389/fncom.2020.00029

Long, N. M. (2023). The intersection of the retrieval state and internal attention. *Nature Communications*, *14*(1), 3861. https://doi.org/10.1038/s41467-023-39609-9

Long, N. M., & Kuhl, B. A. (2019). Decoding the tradeoff between encoding and retrieval to predict memory for overlapping events. *NeuroImage*, *201*, 116001. https://doi.org/10.1016/j.neuroimage.2019.07.014

Long, N. M., Kuhl, B. A., & Chun, M. M. (2018). Memory and Attention. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–37). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119170174.epcn109

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, *62*, 100–108. https://doi.org/10.1016/j.neubiorev.2016.01.003

Mackworth, N. H. (1948). The Breakdown of Vigilance during Prolonged Visual Search1. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21. https://doi.org/10.1080/17470214808416738

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), 87–91. https://doi.org/10.1038/s41586-020-2870-z

Madore, K. P., & Wagner, A. D. (2022). Readiness to remember: Predicting variability in episodic memory. *Trends in Cognitive Sciences*, *26*(8), 707–723. https://doi.org/10.1016/j.tics.2022.05.006

Manning, J. R., Kahana, M. J., & Norman, K. A. (2014). The role of context in episodic memory. In *The cognitive neurosciences, 5th ed* (pp. 557–565). Boston Review. https://doi.org/10.7551/mitpress/9504.003.0061

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences*, *108*(31), 12893–12897. https://doi.org/10.1073/pnas.1015174108

Manza, P., Tomasi, D., Demiral, Ş. B., Shokri-Kojori, E., Lildharrie, C., Lin, E., Wang, G.-J., & Volkow, N. D. (2025). Neural basis for individual differences in the attention-enhancing effects of methylphenidate. *Proceedings of the National Academy of Sciences*, *122*(13), e2423785122. https://doi.org/10.1073/pnas.2423785122

Mazaheri, A., Nieuwenhuis, I. L. C., van Dijk, H., & Jensen, O. (2009). Prestimulus alpha and mu activity predicts failure to inhibit motor responses. *Human Brain Mapping*, *30*(6), 1791–1800. https://doi.org/10.1002/hbm.20763

Mennen, A. C., Turk-Browne, N. B., Wallace, G., Seok, D., Jaganjac, A., Stock, J., deBettencourt, M. T., Cohen, J. D., Norman, K. A., & Sheline, Y. I. (2021). Cloud-Based Functional Magnetic Resonance Imaging Neurofeedback to Reduce the Negative Attentional Bias in Depression: A Proof-of-Concept Study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(4), 490–497. https://doi.org/10.1016/j.bpsc.2020.10.006

Merkow, M. B., Burke, J. F., Stein, J. M., & Kahana, M. J. (2014). Prestimulus theta in the human hippocampus predicts subsequent recognition but not recall. *Hippocampus*, *24*(12), 1562–1569. https://doi.org/10.1002/hipo.22335

Mirjalili, S., & Duarte, A. (2025). Using machine learning to simultaneously quantify multiple cognitive components of episodic memory. *Nature Communications*, *16*(1), 2856. https://doi.org/10.1038/s41467-025-58265-9

Murray, A. M., Nobre, A. C., & Stokes, M. G. (2011). Markers of preparatory attention predict visual short-term memory performance. *Neuropsychologia*, *49*(6), 1458–1465. https://doi.org/10.1016/j.neuropsychologia.2011.02.016

Myers, N. E., Stokes, M. G., Walther, L., & Nobre, A. C. (2014). Oscillatory Brain State Predicts Variability in Working Memory. *Journal of Neuroscience*, *34*(23), 7735–7743. https://doi.org/10.1523/JNEUROSCI.4741-13.2014

Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.58

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal of Neuroscience*, *29*(26), 8604–8611. https://doi.org/10.1523/JNEUROSCI.5967-08.2009

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9*(4), 489–491. https://doi.org/10.1038/nn1663

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601–605. https://doi.org/10.1038/s41586-021-03390-w

Patil, A., & Duncan, K. (2018). Lingering Cognitive States Shape Fundamental Mnemonic Abilities. *Psychological Science*, *29*(1), 45–55. https://doi.org/10.1177/0956797617728592

Peng, K., Wammes, J. D., Nguyen, A., Iordan, C. R., Norman, K. A., & Turk-Browne, N. B. (2024). Inducing representational change in the hippocampus through real-time neurofeedback. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *379*(1915), 20230091. https://doi.org/10.1098/rstb.2023.0091

Pessoa, L., Gutierrez, E., Bandettini, P. A., & Ungerleider, L. G. (2002). Neural Correlates of Visual Working Memory: fMRI Amplitude Predicts Task Performance. *Neuron*, *35*(5), 975–987. https://doi.org/10.1016/S0896-6273(02)00817-6

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., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science (New York, N.Y.)*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Rahmati, M., Saber, G. T., & Curtis, C. E. (2018). Population Dynamics of Early Visual Cortex during Working Memory. *Journal of Cognitive Neuroscience*, *30*(2), 219–233. https://doi.org/10.1162/jocn_a_01196

Riddle, J., Scimeca, J. M., Cellier, D., Dhanani, S., & D'Esposito, M. (2020). Causal Evidence for a Role of Theta and Alpha Oscillations in the Control of Working Memory. *Current Biology: CB*, *30*(9), 1748-1754.e4. https://doi.org/10.1016/j.cub.2020.02.065

Roberts, B. R. T., Pruin, J., Bainbridge, W. A., Rosenberg, M. D., & deBettencourt, M. T. (2025). Memory augmentation with an adaptive cognitive interface. *Psychonomic Bulletin & Review*, *32*(2), 875–886. https://doi.org/10.3758/s13423-024-02589-y

Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, *81*(2), 407–419. https://doi.org/10.3758/s13414-018-1618-4

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, *20*(1), 3–18. https://doi.org/10.1037/0882-7974.20.1.3

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., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., Ramani, R., Qiu, M., Constable, R. T., & Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, *117*(7), 3797–3807. https://doi.org/10.1073/pnas.1912226117

Rosenberg, M. D., Zhang, S., Hsu, W.-T., Scheinost, D., Finn, E. S., Shen, X., Constable, R. T., Li, C.-S. R., & Chun, M. M. (2016). Methylphenidate Modulates Functional Network Connectivity to Enhance Attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(37), 9547–9557. https://doi.org/10.1523/JNEUROSCI.1746-16.2016

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics*, *75*(3), 426–439. https://doi.org/10.3758/s13414-012-0413-x

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

Rudoler, J. H., Bruska, J. P., Chang, W., Dougherty, M. R., Katerman, B. S., Halpern, D. J., Diamond, N. B., & Kahana, M. J. (2024). Decoding EEG for optimizing naturalistic memory. *Journal of Neuroscience Methods*, *410*, 110220. https://doi.org/10.1016/j.jneumeth.2024.110220

Salari, N., & Rose, M. (2016). Dissociation of the functional relevance of different pre-stimulus oscillatory activity for memory formation. *NeuroImage*, *125*, 1013–1021. https://doi.org/10.1016/j.neuroimage.2015.10.037

Scheinost, D., Hsu, T. W., Avery, E. W., Hampson, M., Constable, R. T., Chun, M. M., & Rosenberg, M. D. (2020). Connectome-based neurofeedback: A pilot study to improve sustained attention. *NeuroImage*, *212*, 116684. https://doi.org/10.1016/j.neuroimage.2020.116684

Seeburger, D. T., Xu, N., Ma, M., Larson, S., Godwin, C., Keilholz, S. D., & Schumacher, E. H. (2024). Time-varying functional connectivity predicts fluctuations in sustained attention in a serial tapping task. *Cognitive, Affective & Behavioral Neuroscience*, *24*(1), 111–125. https://doi.org/10.3758/s13415-024-01156-1

Sheehan, T. C., & Serences, J. T. (2022). Attractive serial dependence overcomes repulsive neuronal adaptation. *PLOS Biology*, *20*(9), e3001711. https://doi.org/10.1371/journal.pbio.3001711

Shelat, S., Schooler, J. W., & Giesbrecht, B. (2024). Predicting attentional lapses using response time speed in continuous performance tasks. *Frontiers in Cognition*, *3*. https://doi.org/10.3389/fcogn.2024.1460349

Sheremata, S. L., Somers, D. C., & Shomstein, S. (2018). Visual Short-Term Memory Activity in Parietal Lobe Reflects Cognitive Processes beyond Attentional Selection. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(6), 1511–1519. https://doi.org/10.1523/JNEUROSCI.1716-17.2017

Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, *13*(1), 106–127. https://doi.org/10.1111/tops.12505

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M. L., Rana, M., Oblak, E., Birbaumer, N., & Sulzer, J. (2017). Closed-loop brain

training: The science of neurofeedback. *Nature Reviews. Neuroscience*, *18*(2), 86–100. https://doi.org/10.1038/nrn.2016.164
Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proceedings of the National Academy of Sciences*, *118*(33), e2021905118. https://doi.org/10.1073/pnas.2021905118
Song, H., & Rosenberg, M. D. (2021). Predicting attention across time and contexts with functional brain connectivity. *Current Opinion in Behavioral Sciences*, *40*, 33–44. https://doi.org/10.1016/j.cobeha.2020.12.007
Song, H., Shim, W. M., & Rosenberg, M. D. (2023). Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *eLife*, *12*, e85487. https://doi.org/10.7554/eLife.85487
Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5
Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention, Perception & Psychophysics*, *79*(7), 1897–1916. https://doi.org/10.3758/s13414-017-1357-y
Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron*, *91*(3), 694–707. https://doi.org/10.1016/j.neuron.2016.07.006
St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(23), 6186–6192. https://doi.org/10.1523/JNEUROSCI.4390-15.2016
Stoeckel, L. E., Garrison, K. A., Ghosh, S., Wighton, P., Hanlon, C. A., Gilman, J. M., Greer, S., Turk-Browne, N. B., deBettencourt, M. T., Scheinost, D., Craddock, C., Thompson, T., Calderon, V., Bauer, C. C., George, M., Breiter, H. C., Whitfield-Gabrieli, S., Gabrieli, J. D., LaConte, S. M., … Evins, A. E. (2014). Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *NeuroImage : Clinical*, *5*, 245–255. https://doi.org/10.1016/j.nicl.2014.07.002
Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004
Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M. L., Bruehl, A. B., Cohen, L. G., DeCharms, R. C., Gassert, R., Goebel, R., Herwig, U., LaConte, S., Linden, D., Luft, A., Seifritz, E., & Sitaram, R. (2013). Real-time fMRI neurofeedback: Progress and challenges. *NeuroImage*, *76*, 386–399. https://doi.org/10.1016/j.neuroimage.2013.03.033
Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K., & Awh, E. (2019). Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of Neurophysiology*, *122*(2), 539–551. https://doi.org/10.1152/jn.00268.2019
Swallow, K. M., & Jiang, Y. V. (2010). The Attentional Boost Effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*(1), 118–132. https://doi.org/10.1016/j.cognition.2009.12.003
Swallow, K. M., & Jiang, Y. V. (2013). Attentional Load and Attentional Boost: A Review of Data and Theory. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00274

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

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., Kawato, M., & Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3470–3475. https://doi.org/10.1073/pnas.1721572115

Terashima, H., Kihara, K., Kawahara, J. I., & Kondo, H. M. (2021). Common principles underlie the fluctuation of auditory and visual sustained attention. *Quarterly Journal of Experimental Psychology (2006)*, *74*(4), 705–715. https://doi.org/10.1177/1747021820972255

Trach, J. E., deBettencourt, M. T., Radulescu, A., & McDougle, S. D. (2025). Rewards transiently and automatically enhance sustained attention. *Journal of Experimental Psychology: General*, *154*(4), 1063–1079. https://doi.org/10.1037/xge0001727

Tran, T. T., Madore, K. P., Tobin, K. E., Block, S. H., Puliyadi, V., Hsu, S. C., Preston, A. R., Bakker, A., & Wagner, A. D. (2025). Age-Related differences in the relationship between sustained attention and associative memory and Memory-Guided inference. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-025-01292-2

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

Turk-Browne, N. B., Yi, D.-J., & Chun, M. M. (2006). Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron*, *49*(6), 917–927. https://doi.org/10.1016/j.neuron.2006.01.030

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

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*(1), 111–122. https://doi.org/10.1016/j.intell.2009.08.002

Unsworth, N., & Robison, M. K. (2016a). 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. (2016b). The influence of lapses of attention on working memory capacity. *Memory & Cognition*, *44*(2), 188–196. https://doi.org/10.3758/s13421-015-0560-0

Unsworth, N., & Robison, M. K. (2020). Working memory capacity and sustained attention: A cognitive-energetic perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(1), 77–103. https://doi.org/10.1037/xlm0000712

van Dijk, H., Schoffelen, J.-M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(8), 1816–1823. https://doi.org/10.1523/JNEUROSCI.1853-07.2008

van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477–483. https://doi.org/10.1038/s41593-018-0335-6

van Ede, F., Niklaus, M., & Nobre, A. C. (2017). Temporal Expectations Guide Dynamic Prioritization in Visual Working Memory through Attenuated α Oscillations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(2), 437–445. https://doi.org/10.1523/JNEUROSCI.2272-16.2016

Vo, V. A., Sutterer, D. W., Foster, J. J., Sprague, T. C., Awh, E., & Serences, J. T. (2022). Shared Representational Formats for Information Maintained in Working Memory and Information Retrieved from Long-Term Memory. *Cerebral Cortex*, *32*(5), 1077–1092. https://doi.org/10.1093/cercor/bhab267

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

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

Wan, Q., Ardalan, A., Fulvio, J. M., & Postle, B. R. (2024). Representing Context and Priority in Working Memory. *Journal of Cognitive Neuroscience*, *36*(7), 1374–1394. https://doi.org/10.1162/jocn_a_02166

Wang, Y. C., & Egner, T. (2023). Target detection does not influence temporal memory. *Attention, Perception, & Psychophysics*, *85*(6), 1936–1948. https://doi.org/10.3758/s13414-023-02723-3

Weidemann, C. T., & Kahana, M. J. (2021). Neural measures of subsequent memory reflect endogenous variability in cognitive function. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *47*(4), 641–651. https://doi.org/10.1037/xlm0000966

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), 971–978. https://doi.org/10.1038/nn1727

Yamashita, A., Hayasaka, S., Kawato, M., & Imamizu, H. (2017). Connectivity Neurofeedback Training Can Differentially Change Functional Connectivity and Cognitive Performance. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(10), 4960–4970. https://doi.org/10.1093/cercor/bhx177

Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M., & Esterman, M. (2021). Brain state-based detection of attentional fluctuations and their modulation. *NeuroImage*, *236*, 118072. https://doi.org/10.1016/j.neuroimage.2021.118072

Yeh, N., & Rose, N. S. (2019). How Can Transcranial Magnetic Stimulation Be Used to Modulate Episodic Memory?: A Systematic Review and Meta-Analysis. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00993

Yoo, J. J., Hinds, O., Ofen, N., Thompson, T. W., Whitfield-Gabrieli, S., Triantafyllou, C., & Gabrieli, J. D. E. (2012). When the brain is prepared to learn: Enhancing human learning using real-time fMRI. *NeuroImage*, *59*(1), 846–852. https://doi.org/10.1016/j.neuroimage.2011.07.063

Yoo, K., Rosenberg, M. D., Kwon, Y. H., Lin, Q., Avery, E. W., Sheinost, D., Constable, R. T., & Chun, M. M. (2022). A brain-based general measure of attention. *Nature Human Behaviour*, *6*(6), 782–795. https://doi.org/10.1038/s41562-022-01301-1

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

Zhang, Z., & Lewis-Peacock, J. A. (2024). Integration of history information Drives Serial Dependence and Stabilizes Working Memory Representations. *Journal of Neuroscience*, *44*(32). https://doi.org/10.1523/JNEUROSCI.2399-23.2024

Zhang, Z., & Rosenberg, M. D. (2024). Assessing the impact of attention fluctuations on statistical learning. *Attention, Perception, & Psychophysics*, *86*(4), 1086–1107. https://doi.org/10.3758/s13414-023-02805-2

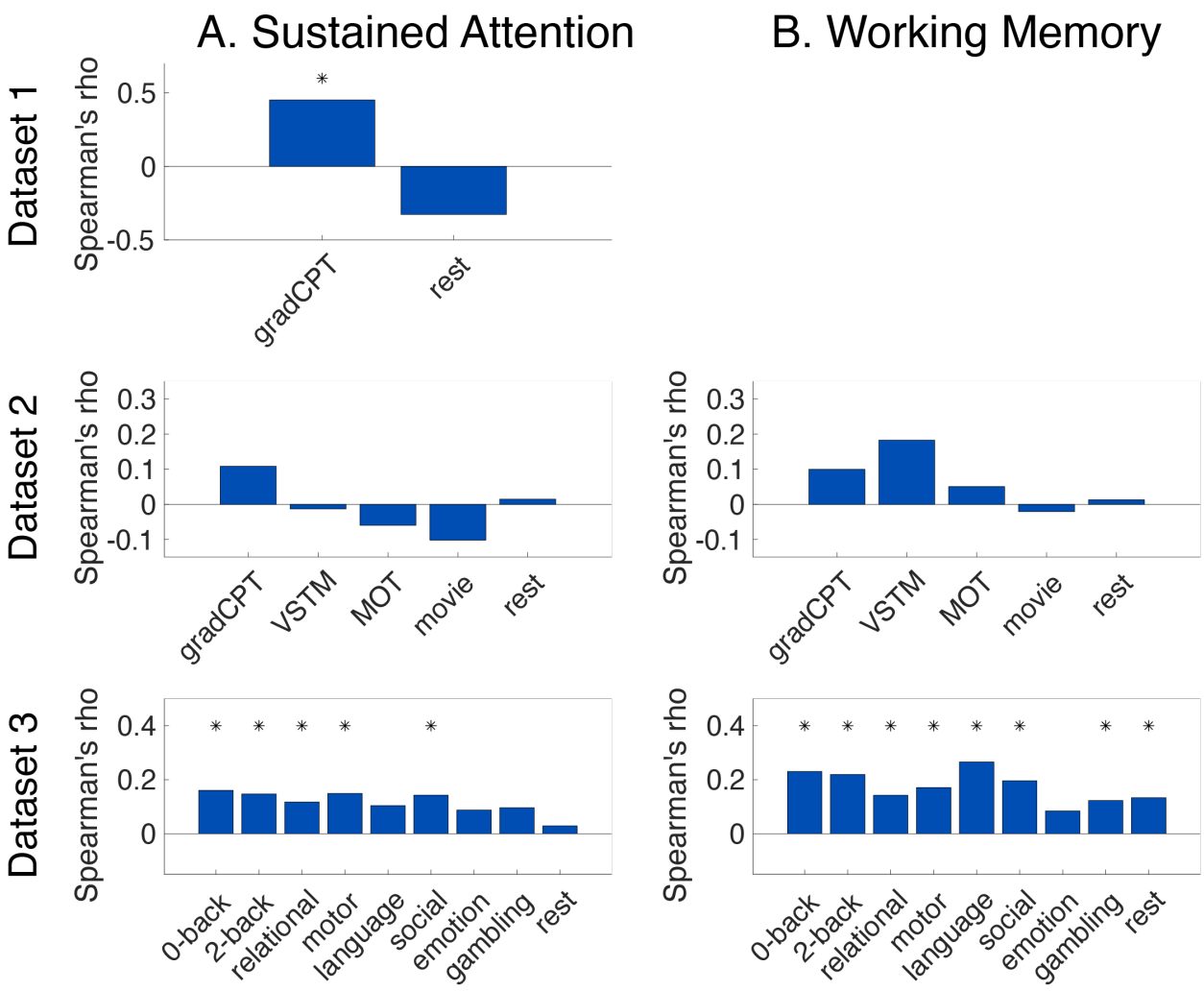
Zhao, C., Corriveau, A., Ke, J., Vogel, E. K., & Rosenberg, M. D. (2025). *Sustained attention is more closely related to long-term memory than to attentional control* (p. 2025.03.13.643171). bioRxiv. https://doi.org/10.1101/2025.03.13.643171

Zhao, C., & Woodman, G. F. (2021). Converging Evidence That Neural Plasticity Underlies Transcranial Direct-Current Stimulation. *Journal of Cognitive Neuroscience*, *33*(1), 146–157. https://doi.org/10.1162/jocn_a_01639

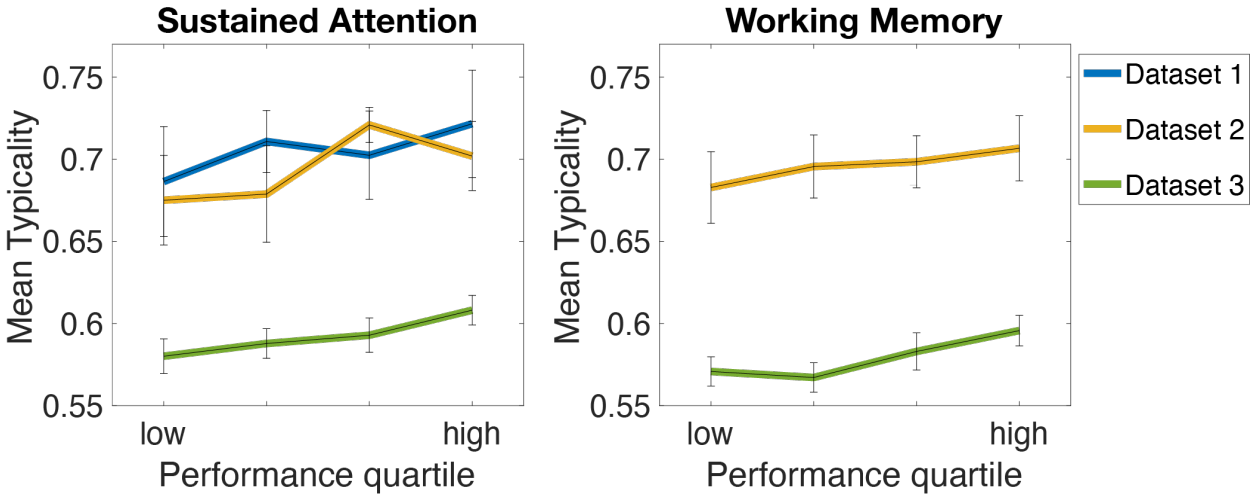
Zokaei, N., Board, A. G., Manohar, S. G., & Nobre, A. C. (2019). Modulation of the pupillary response by the content of visual working memory. *Proceedings of the National Academy of Sciences*, *116*(45), 22802–22810. https://doi.org/10.1073/pnas.1909959116

Zuberer, A., Kucyi, A., Yamashita, A., Wu, C. M., Walter, M., Valera, E. M., & Esterman, M. (2021). Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *NeuroImage*, *229*, 117610. https://doi.org/10.1016/j.neuroimage.2020.117610

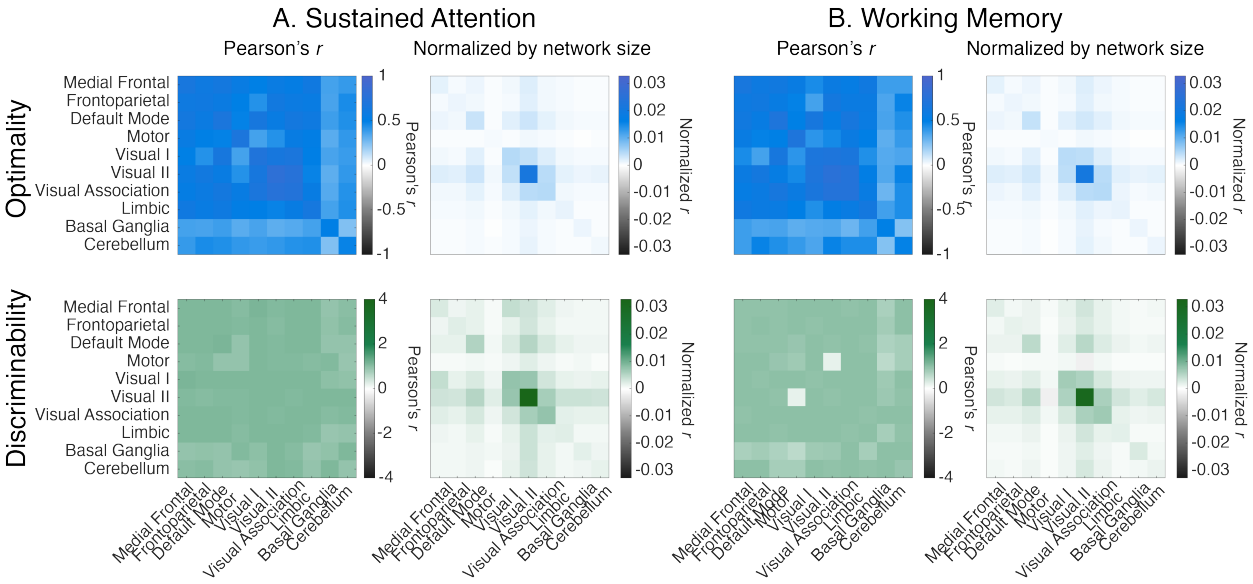
## Appendix 1: Chapter 1 supplementary materials




*Appendix 1 Figure 36. Correlations between connectome features and performance.* Correlations between averaged within-run typicality and performance on (A.) sustained attention and (B.) working memory tasks. Bars indicate the partial Spearman correlation between pairwise run typicality and performance, controlling for head motion and within and between runs. Significance $p < .05$ (uncorrected) is indicated with asterisks.

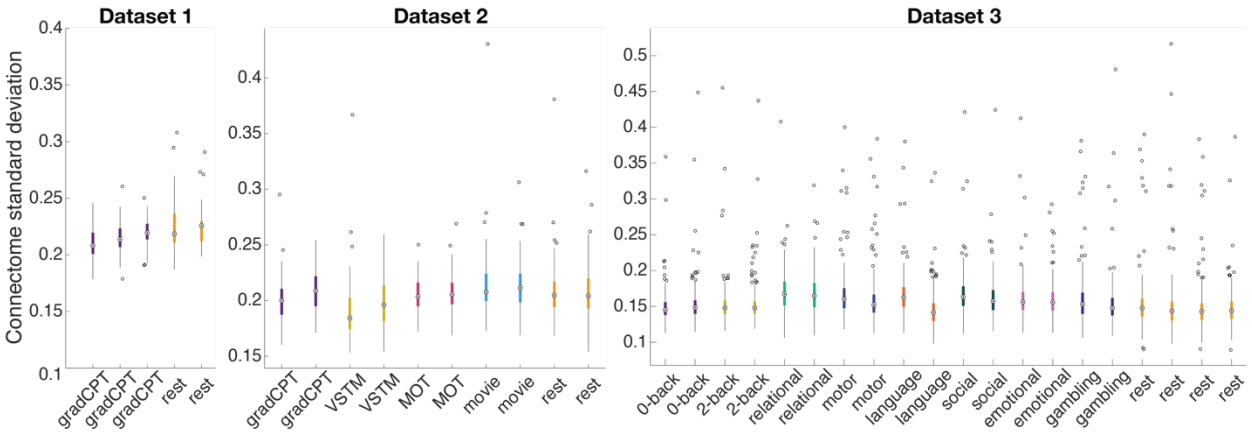



*Appendix 1 Figure 37. Mean typicality generally increases across performance quartiles*. Error bars indicate standard error for each quartile.

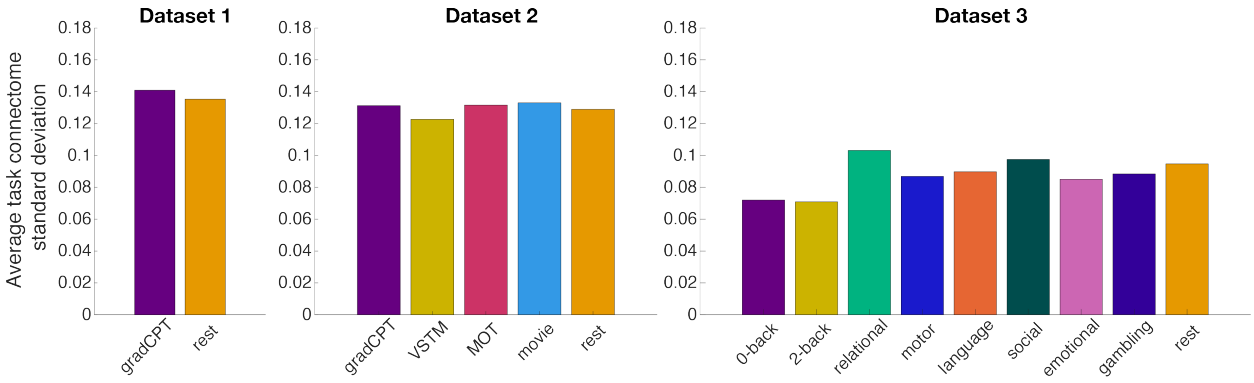



*Appendix 1 Figure 38. Network connections for connectome optimality and discriminability.* Within- and between-network connections for (A.) sustained attention and (B.) working memory task connectomes for the connectome features optimality and discriminability. Matrices on the left represent average values (Pearson's correlations for optimality matrices, ratio of Pearson's

correlations for discriminability matrices) across datasets. Matrices on the right are normalized by network size.




*Appendix 1 Figure 39. Within-connectome standard deviation values across participants.*




*Appendix 1 Figure 40. Group- and task-averaged connectome standard deviation values.*

| | **Shen parcellation (268 nodes)** | | **Yeo parcellation (122 nodes)** | | **Steiger's Z test of difference** | |
|---|---|---|---|---|---|---|
| | Stability-sustained attention correlation | *p*-value | Stability-sustained attention correlation | *p*-value | Steiger's Z | *p*-value |
| Stability | .344 | .109 | .520 | .011 | 1.056 | .302 |
| Typicality | .452 | .030 | .527 | 9.75*10-3 | .757 | .457 |
| Optimality | .801 | $1.27*10^{-5}$ | .608 | 3.44*10-3 | -1.619 | .121 |
| Discriminability | .109 | .622 | .214 | .326 | -.728 | .474 |

*Appendix 1 Table 6. Comparison of prediction by parcellation scheme.* Prediction of sustained attention performance from individual and group-level connectome features calculated using 268-node (Shen et al., 2013) and 122-node (Yeo et al., 2011) parcellation schemes. Correlation values were not significantly different between parcellation schemes.

| | | Stability | | Typicality | | Optimality | | Discriminability | |
|---|---|---|---|---|---|---|---|---|---|
| | | rho | sig. | rho | sig. | rho | sig. | rho | sig. |
| Dataset 1 | Sustained Attention | .344 | .109 | .452 | .030 * | .801 | $1.27*10^{-5}$ *** | .109 | .622 |
| Dataset 2 | Sustained Attention | .411 | $8.22*10^{-4}$ *** | .108 | .397 | .445 | $3.63*10^{-4}$. *** | .415 | $7.20*10^{-4}$ *** |
| | Working Memory | .264 | .027 * | .183 | .130 | .258 | .036 * | .156 | .197 |
| Dataset 3 | Sustained Attention | .168 | $3.34*10^{-3}$ ** | .161 | $5.17*10^{-3}$ ** | .235 | $1.48*10^{-4}$ *** | .128 | .027 * |
| | Working Memory | .207 | $2.54*10^{-4}$ *** | .219 | $1.03*10^{-4}$ *** | .249 | $1.75*10^{-5}$ *** | .150 | $8.31*10^{-3}$ ** |

*Appendix 1 Table 7. Across-subject Spearman's rank correlation values between connectome features and task performance.* Significance values are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

| | | | Testing Dataset | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Dataset 1 | | Dataset 2 | | Dataset 3 | |
| | | | rho | sig. | rho | sig. | rho | sig. |
| Optimal Template Dataset | Sustained Attention | Dataset 1 | .801 | $1.27*10^{-5}$ *** | $-2.62*10^{-4}$ | .998 | .179 | $1.83*10^{-3}$ ** |
| | | Dataset 2 | .542 | $7.55*10^{-3}$ ** | .445 | $3.63*10^{-4}$. *** | .094 | .102 |
| | | Dataset 3 | .485 | .019 * | -.066 | .605 | .258 | .036 * |
| | Working Memory | Dataset 2 | | | .235 | $1.48*10^{-4}$ *** | .078 | .173 |
| | | Dataset 3 | | | .194 | .108 | .249 | $1.75*10^{-5}$ *** |

*Appendix 1 Table 8. Optimality generalizes to predict sustained attention*. Connectome optimality or similarity to top performers' mean connectome generalizes across dataset to predict performance in sustained attention tasks but not working memory tasks. Significance values are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

| | | Single top score | | | | Top 10% of performers | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | N | Optimality-performance correlation | Significance | | N | Optimality-performance correlation | Significance | |
| DS1 | Sustained attention | 1 | .770 | $2.73*10^{-5}$ | *** | 2 | .783 | $7.36*10^{-5}$ | *** |
| DS2 | Sustained attention | 1 | .401 | $1.13*10^{-3}$ | ** | 10 | .473 | $2.35*10^{-4}$ | *** |
| | Working memory | 1 | .245 | .042 | * | 10 | .246 | .054 | |
| DS3 | Sustained attention | 46 | .235 | $1.48*10^{-4}$ | *** | 46 | .235 | $1.48*10^{-4}$ | *** |
| | Working memory | 3 | .276 | $9.54*10^{-7}$ | *** | 34 | .210 | $4.41*10^{-4}$ | *** |

*Appendix 1 Table 9***.** *Comparisons of optimality thresholds*. Predictions of sustained attention and working memory behavioral performance from similarity to the optimal connectome, calculated as the average functional connectivity pattern of performers who achieved the single highest performance score or of performers who achieved the top 10% of performance scores.

Sustained Attention

| | Dataset 2 | | | | Dataset 3 | | | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Std Error | Significance | AIC | Coefficient | Std Error | Significance | AIC |
| saCPM strength | .5887 | .1026 | $3.68*10^{-7}$ *** | **140.8** | .0854 | .0596 | .1533 | 719.4 |
| Stability | .3419 | .1089 | $2.67*10^{-3}$ ** | 158.9 | .1760 | .0639 | $6.34*10^{-3}$ ** | 713.9 |
| saCPM strength | .5340 | .1238 | 6.63e-05 *** | | .1159 | .0675 | .0874 | |
| Stability | .1574 | .1663 | .348 | 143.8 | .1593 | .0648 | .0147 * | 713.5 |
| saCPM strength * stability | .0466 | .0683 | .498 | | -.1147 | .0618 | .0647 | |
| Typicality | .2205 | .1984 | .271 | 167.3 | .1982 | .0605 | $1.21*10^{-3}$ ** | 710.8 |
| saCPM strength | .6425 | .1146 | $6.56*10^{-7}$ *** | | .0691 | .0600 | .2502 | |
| Typicality | -.2052 | .1896 | .284 | 143.5 | .1464 | .0629 | .0208 * | 705.7 |
| saCPM strength * typicality | -.0583 | .1170 | .620 | | -.1839 | .0632 | $3.91*10^{-3}$ ** | |
| Optimality | .5955 | .1582 | $3.92*10^{-4}$ *** | 155.1 | .2384 | .0601 | $9.53*10^{-5}$ *** | 706.0 |
| saCPM strength | .5384 | .1395 | $2.97*10^{-4}$ *** | | .0706 | .0593 | .2341 | |
| Optimality | .1166 | .1959 | .554 | 144.3 | .1890 | .0619 | $2.53*10^{-3}$ ** | **700.2** |
| saCPM strength * optimality | .0370 | .1044 | .725 | | -.1933 | .0636 | $2.61*10^{-3}$ ** | |

Working Memory

| | Dataset 2 | | | |
|---|---|---|---|---|
| | Coefficient | Std Error | Significance | AIC |
| WM CPM strength | -.0380 | .1147 | .742 | 191.1 |
| Stability | .4060 | .1147 | $7.54*10^{-4}$ *** | 179.1 |
| WM CPM strength | .0757 | .1108 | .497 | |
| Stability | . 4415 | .1114 | $1.94*10^{-4}$ *** | **175.9** |
| WM CPM strength * stability | -.3064 | .1183 | .012 * | |
| Typicality | .3256 | .1154 | $6.36*10^{-3}$ ** | 183.3 |
| WM CPM strength | -.1626 | .1169 | .169 | |
| Typicality | .3906 | .1250 | $2.70*10^{-3}$ ** | 185.1 |
| WM CPM strength * typicality | .0600 | .1210 | .622 | |
| Optimality | .3141 | .1137 | $7.49*10^{-3}$ ** | 183.6 |
| WM CPM strength | -.1375 | .1150 | .2361 | |
| Optimality | .3614 | .1217 | $4.25*10^{-3}$ ** | 186.0 |
| WM CPM strength * optimality | .0439 | .1227 | .7217 | |

*Appendix 1 Table 10. Contributions of network-based predictive models and connectome strength towards behavior prediction.* Note- sustained attention models were not tested in Dataset 1 and working memory models were not tested in Dataset 3 because the network-based predictive models

were originally trained in these datasets. Bolded AIC values indicate the best model for predicting the sustained attention and working memory performance. Significance values are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

| | Run type | # TRs | Mean Stability | Stability-sustained attention correlation | uncorrected *p*-value | | Stability-working memory performance correlation | uncorrected *p*-value | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset 1 | gradCPT | 816 | 0.7611 | 0.3170 | 0.1615 | | NA | NA | |
| | gradCPT | 816 | 0.7511 | 0.5472 | 0.0230 | * | NA | NA | |
| | gradCPT | 816 | 0.7788 | 0.4198 | 0.0934 | | NA | NA | |
| | rest | 360 | 0.5700 | 0.2598 | 0.2687 | | NA | NA | |
| Dataset 2 | gradCPT | 600 | 0.7272 | 0.4109 | $8.217*10^{-4}$ | *** | 0.2718 | 0.0311 | * |
| | VSTM | 500 | 0.7040 | -0.0630 | 0.6043 | | 0.2644 | 0.0270 | * |
| | MOT | 600 | 0.7354 | 0.2973 | 0.0146 | * | 0.3502 | 0.0037 | ** |
| | movie | 433 | 0.6639 | 0.1672 | 0.1489 | | 0.0721 | 0.5359 | |
| | rest | 600 | 0.7699 | 0.1499 | 0.1683 | | -0.0876 | 0.4224 | |
| Dataset 3 | 0-back | 200 | 0.3594 | 0.1683 | 0.0033 | ** | 0.2191 | $1.03*10^{-4}$ | *** |
| | 2-back | 200 | 0.3474 | 0.1682 | 0.0034 | ** | 0.2067 | $2.54*10^{-4}$ | *** |
| | relational | 253 | 0.5801 | 0.1009 | 0.0800 | | 0.0911 | 0.1101 | |
| | motor | 284 | 0.4585 | 0.1158 | 0.0443 | * | 0.2358 | $2.82*10^{-5}$ | *** |
| | language | 316 | 0.5770 | 0.1468 | 0.0107 | * | 0.2706 | $1.38*10^{-6}$ | *** |
| | social | 274 | 0.5801 | 0.1405 | 0.0145 | * | 0.2268 | $5.75*10^{-5}$ | *** |
| | emotion | 232 | 0.4378 | 0.0309 | 0.5924 | | 0.0273 | 0.6330 | |
| | gambling | 176 | 0.5405 | 0.0074 | 0.8984 | | 0.0888 | 0.1192 | |
| | rest | 1200 | 0.7058 | -0.0211 | 0.7150 | | 0.1022 | 0.0727 | |
| | rest | 1200 | 0.7163 | 0.0118 | 0.8387 | | 0.1031 | 0.0703 | |
| | rest | 1200 | 0.6642 | 0.0240 | 0.6777 | | 0.1118 | 0.0496 | * |
| | rest | 1200 | 0.6803 | -0.0182 | 0.7525 | | 0.0625 | 0.2735 | |
| | rest | 1200 | 0.7090 | 0.0468 | 0.4181 | | 0.0686 | 0.2291 | |
| | rest | 1200 | 0.6969 | 0.0014 | 0.9809 | | 0.0566 | 0.3215 | |

*Appendix 1 Table 11*. *Scan length, mean task stability, and across-subject Spearman's correlation with behavioral performance scores calculated between pairs of task runs*. Significance values are indicated with asterisks. * $p < .05$ ** $p < .01$ *** $p < .001$

## Appendix 2: Chapter 3 supplementary materials

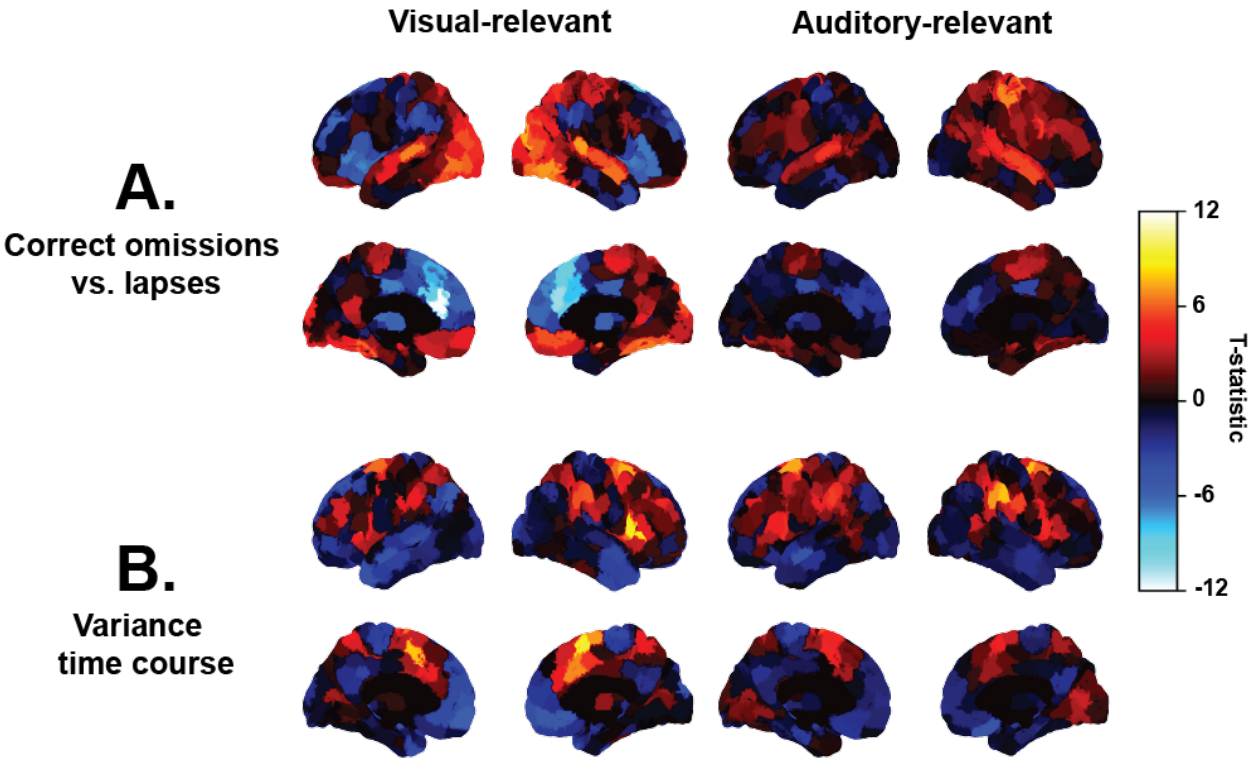
*Simulated edge predictions*

To more closely examine when univariate and edge time series make similar or differing predictions, we simulated univariate time series for two theoretical ROIs ($x_1$, $x_2$) that differed in the extent to which they predicted a theoretical third variable $y$ and in the extent to which they were related to each other. In this case, $y$ serves as our predicted behavior (e.g., attentional state) while univariate time series can be thought of as differing in their functional similarity to one another. We then calculated the extent to which edges between univariate time series predicted behavior ($y$).

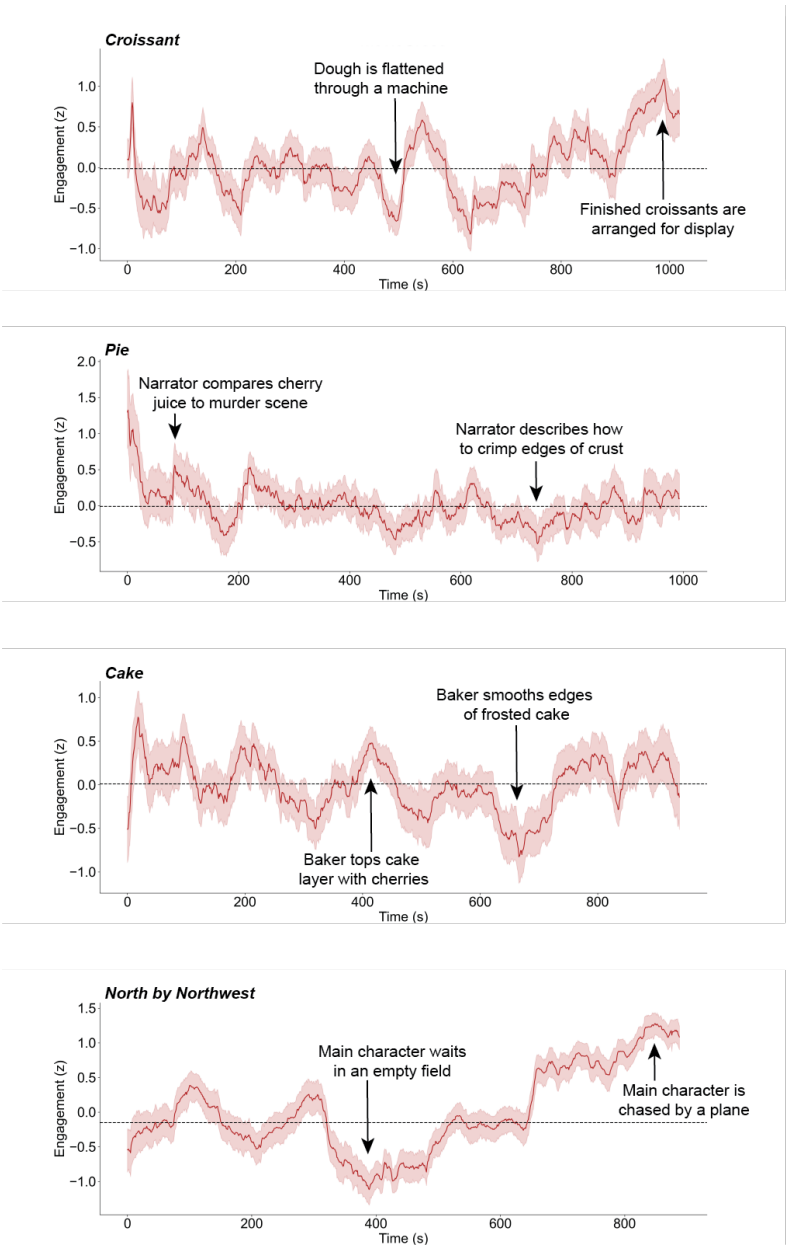
Theoretical variable $y$ time series were generated as a random vector of length 1000 where each value was drawn from a normal distribution between 0 and 1. This time series was smoothed with a gaussian kernel of standard deviation 5. Univariate time series $x_1$ was generated by multiplying $y$ by a randomly-signed value between 0 and 1 (in multiples of 0.1) and adding noise. Noise vectors were generated as a random vector of length 1000, smoothed with a gaussian kernel of standard deviation 5. Finally, to influence the amount of noise applied, noise vectors were multiplied by either 0.1 (low noise), 0.5 (medium noise), or 0.9 (high noise) before being added to $x_1$. Univariate time series $x_2$ was generated by multiplying $x_1$ by a randomly-signed value between 0 and 1 (in multiples of 0.1) and adding a noise vector. This allowed us to manipulate the dependency between $x_1$ and $x_2$. We then calculated the edge between univariate time series as the element-wise product of z-scored time series and correlated this edge time series with $y$. We repeated this process 10,000 times for each dependency between $x_1$ and $y$ and for each dependency between $x_1$ and $x_2$. Analysis code is available at https://osf.io/xk6v9.

We note that, while this simulation method provides some intuition as to when univariate and edge-based dynamics lead to similar relationships with a behavioral variable, it does not capture the full range of possible relationships between neural activity and behavior in this analysis. Further, the assumptions made when generating time series in the current analysis may not reflect the true, observable relationships. For example, there are many observable instances of edge co-fluctuations whose ability to track behavior is not predicted by univariate activity---and vice versa---not well-captured in the current analyses. Future work should seek to characterize the scenarios in which these patterns occur to simulate data that more closely reflects real relationships between the brain and behavior.

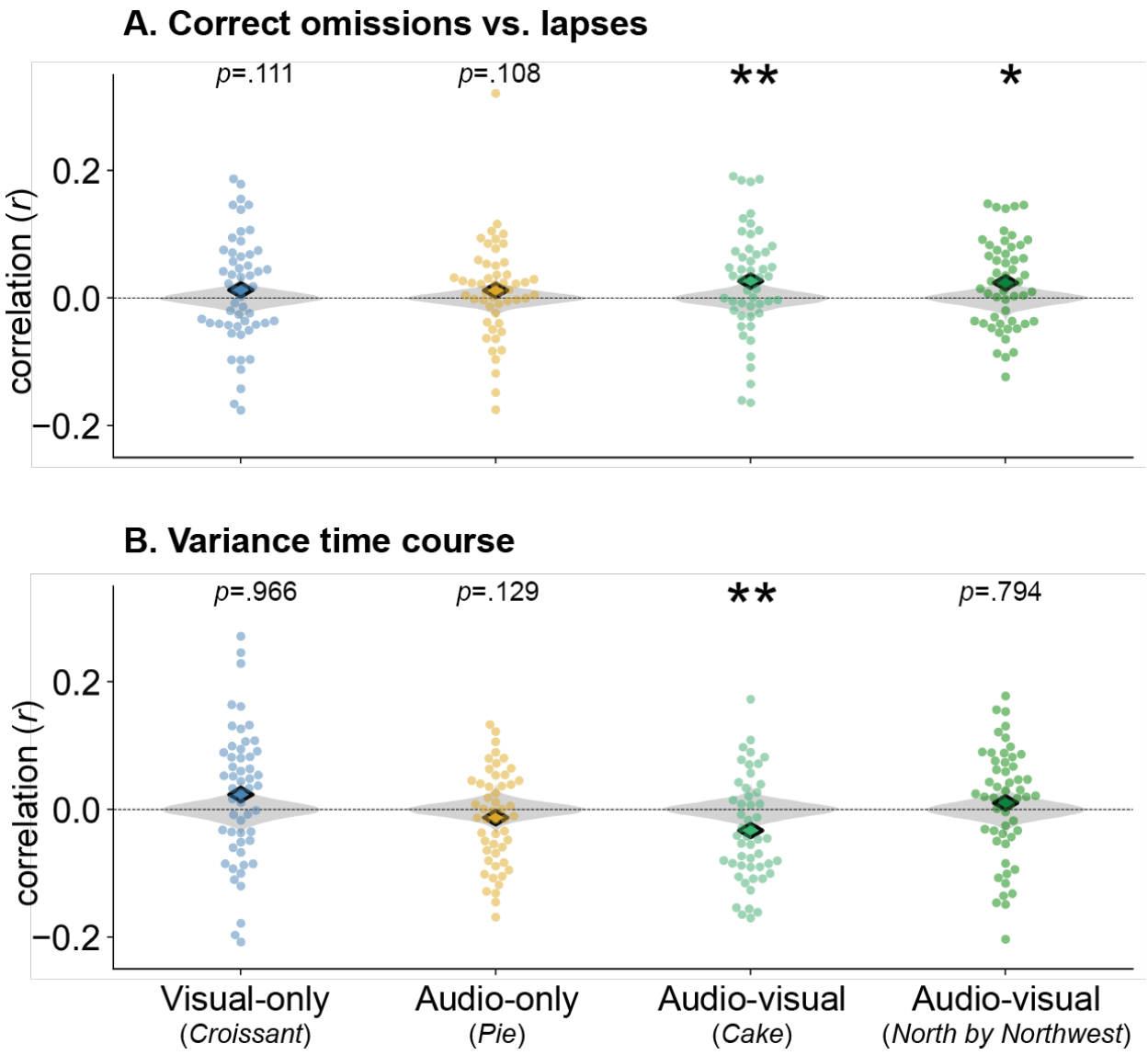
We visualize the observed correlation between edge time series and $y$ as a function of the relationship between both $x_1$ *and* $x_2$ and as a function of the relationship between $x_1$ and $y$ in **Appendix 2 Figure A5a.** We also visualize a random sample of edge predictions as a function of observed correlations between $x_1$, $x_2$, and $y$ to better understand the distribution of observed edge predictions (correlations) based on these relationships (**Appendix 2 Figure A5b**).



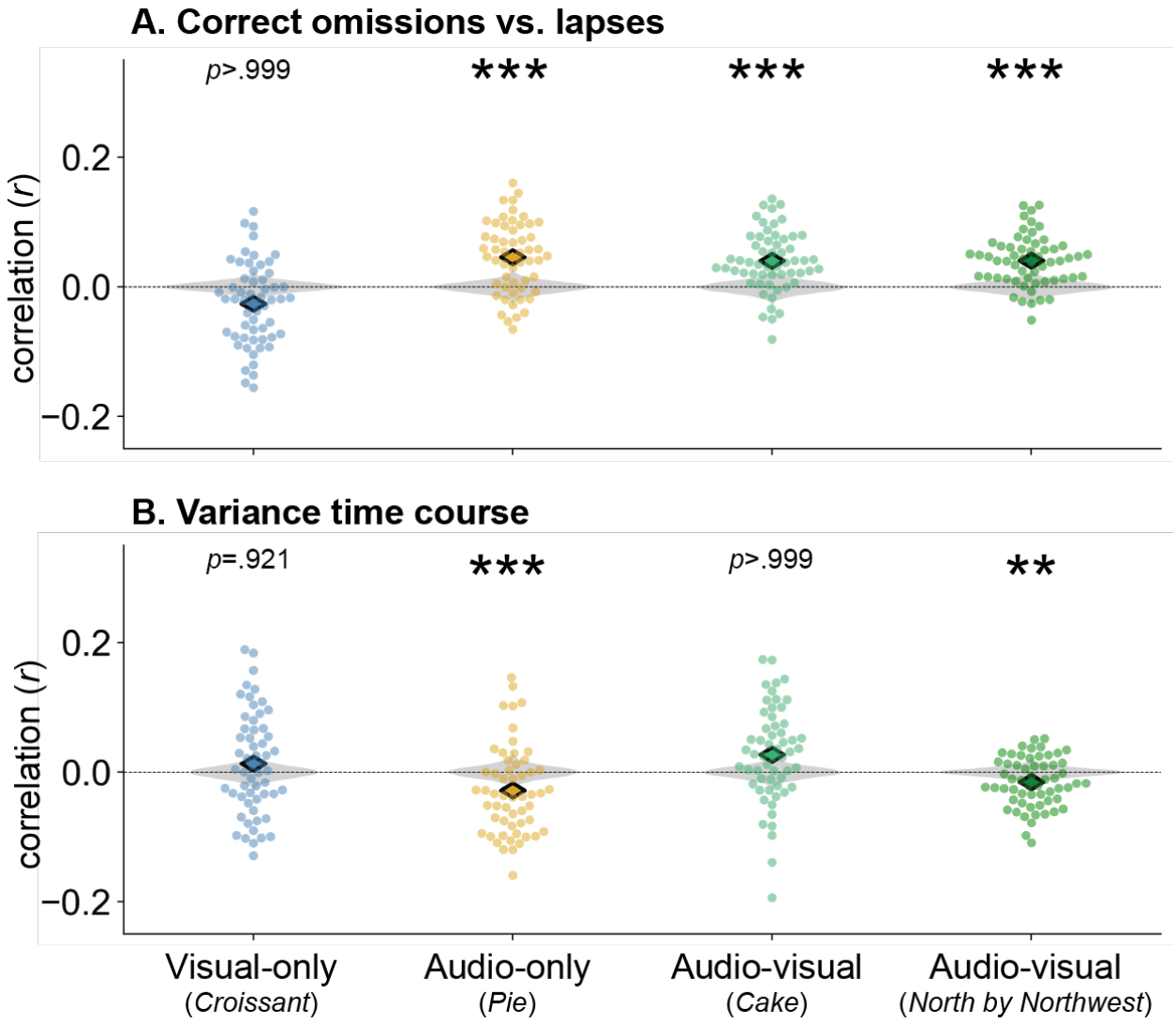

*Appendix 2 Figure 41. Maps visualizing unthresholded results from contrasts of interest*. (A) Activity more strongly evoked by correct inhibition of responses to infrequent trials is visualized in warm colors, whereas activity related to sustained attentional lapses is visualized in cool colors. (B) Univariate results from contrasting the variance time course against baseline. Warm colors indicate areas with increased activity related to greater response time variance, reflecting worse sustained attentional states. Cool colors reflect regions whose activity increased with decreases in response variance, indicating better sustained attention.



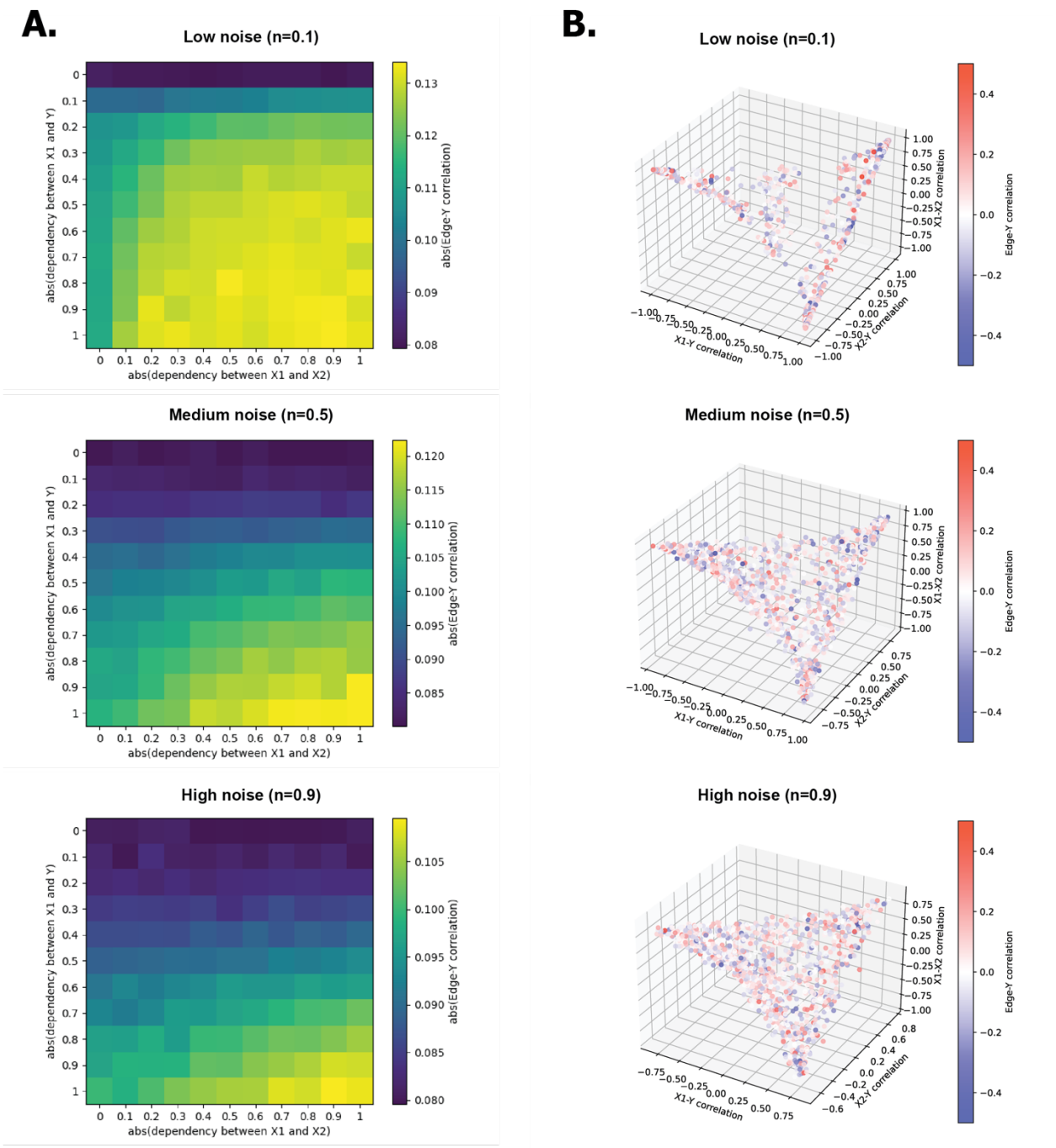

*Appendix 2 Figure 42. Mean engagement time courses for naturalistic narrative stimuli averaged across participants*. Time courses were z-scored within participants before averaging. Shaded regions reflect 95% confidence intervals. Dotted black lines indicate median engagement across the time course.




*Appendix 2 Figure 43. Correlations between subject-level engagement time courses and predicted network strength time courses*. Null distributions generated with circle-shifting are shown in gray. (A) Predictions generated from modality-general edges related to correct omissions vs. lapses were correlated with subject-level engagement in both audio-visual movies. (B) Modality-general edges related to the variance time course predicted subject-level engagement in one audio-visual movie, *Cake*.




*Appendix 2 Figure 44. Correlations between mean engagement time courses and time courses of univariate activity*. Univariate time courses were calculated as the mean activity in regions positively related to sustained attention in both visual-relevant and auditory-relevant runs of the avCPT. Null distributions shown in gray were generated by circle-shifting the predicted univariate time course 10,000 times. (A) Univariate activity in regions related to correct omissions vs. lapses predicted engagement in the podcast and both audio-visual movies. (B) Activity related to the variance time course predicted engagement during the podcast and one audio-visual movie, *North by Northwest.*




*Appendix 2 Figure 45. Visualizing edge predictions as a function of univariate prediction and activity similarity between pairs of ROIs*. A) As univariate activity becomes more strongly predictive of *y,* and as the time series between pairs of regions ($x_1$ and $x_2$) become more related, the edge time series between a pair of regions should also more strongly predict *y.* However, as

noise in ROI time series increases, edges are predictive when an ROI time series strongly predicts *y*, and when it is strongly related to the second ROI time series. B) Edge predictions plotted from a random subset of simulations (10 points from each possible dependency between $x_1$ and $x_2$ as well as between $x_1$ and *y,* i.e., 10 dots per cell from matrix in A). Results demonstrate that, while edges tend to become more predictive as correlations between $x_1$ and $x_2$ as well as between $x_1$ and *y* strengthen, edge prediction is quite variable in general. Edges can be strongly related to *y* (shown as more saturated blue and red points) even if variables $x_1$ and $x_2$ are weakly correlated.

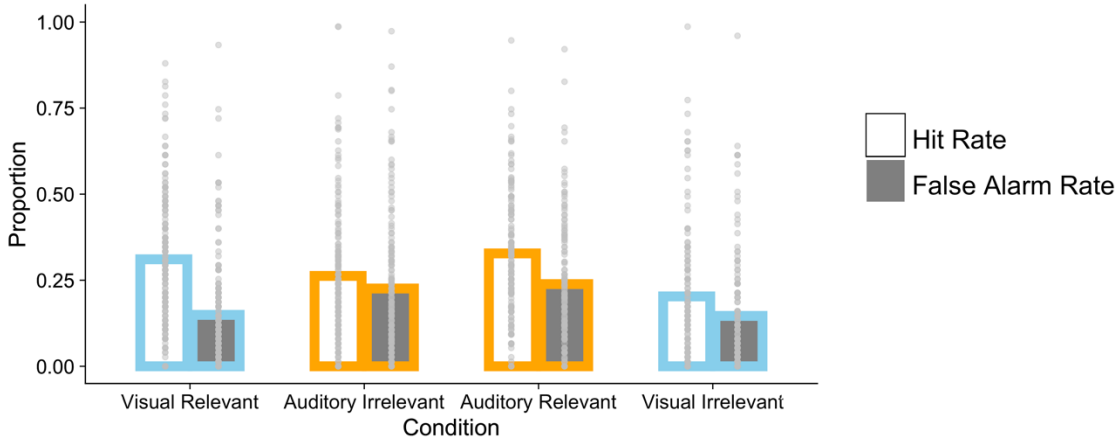
## Appendix 3: Chapter 4 supplementary materials

| | **Experiment 2** | | | **Experiment 3** | | |
|---|---|---|---|---|---|---|
| | coef. | SE | sig. | coef. | SE | sig. |
| Relevance *(irrelevant vs relevant)* | -.504 | .023 | <.001*** | -.931 | .026 | <.001*** |
| Frequency *(infrequent vs frequent)* | -.343 | .022 | <.001*** | -.207 | .027 | <.001*** |
| Category *(scenes vs faces)* | -.584 | .023 | <.001*** | -.214 | .026 | <.001*** |
| RT speed | .010 | .013 | .409 | .010 | .015 | .485 |
| RT variance | .003 | .012 | .813 | -.002 | .013 | .853 |
| RT speed: RT variance | .008 | .010 | .407 | .004 | .012 | .707 |

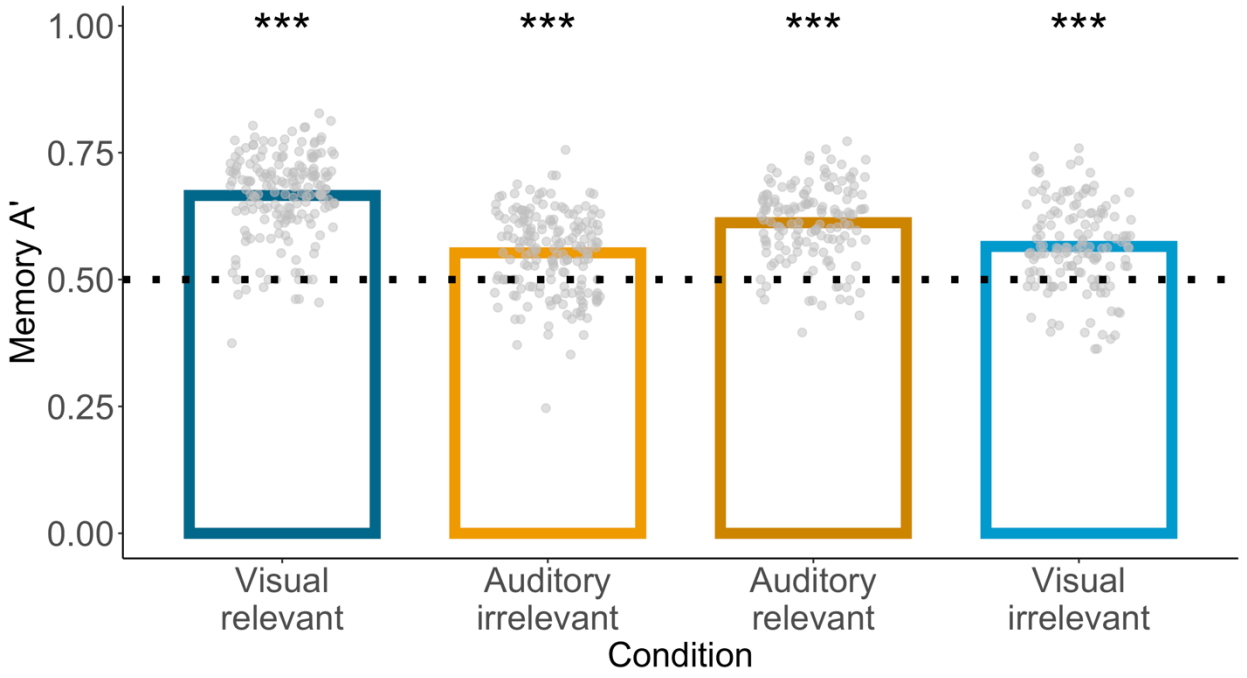
*Appendix 3 Table 12. Results from models predicting stimulus memory from fixed effects of stimulus relevance, frequency, and category, RT speed, RT variance, and their interaction*. Memory judgements were considered accurate if old stimuli were provided a rating greater than 3 (i.e., maybe old or definitely old). Models were fit within-experiment and included a subject-level random intercept term. *** p<.001, ** p<.01, * p<.05

| | Auditory | | | Visual | | |
|---|---|---|---|---|---|---|
| Predictor | Coefficient | Standard Error | Significance | Coefficient | Standard Error | Significance |
| RT speed | .263 | .047 | <.001 *** | .599 | .047 | <.001 *** |
| RT variance | -.173 | .035 | <.001 *** | -.221 | .037 | <.001 *** |
| RT speed:RT variance | -.034 | .030 | .264 | -.041 | .028 | .134 |

*Appendix 3 Table 13. Fixed effects of reaction time measures on sustained attention lapses during the CPT.* Reaction time measures were calculated using only those trials with three presses for correct, frequent preceding trials.****p*<.001, ***p*<.01, **p*<.05

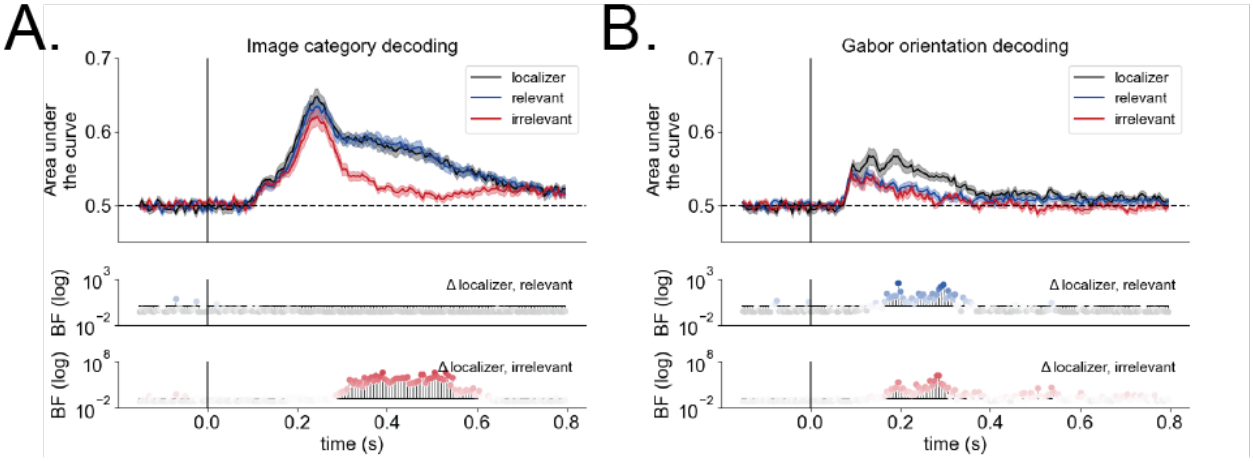



*Appendix 3 Figure 46. Hit and false alarm rates for task-relevant and task-irrelevant memory conditions.* Gray dots represent individual subjects.

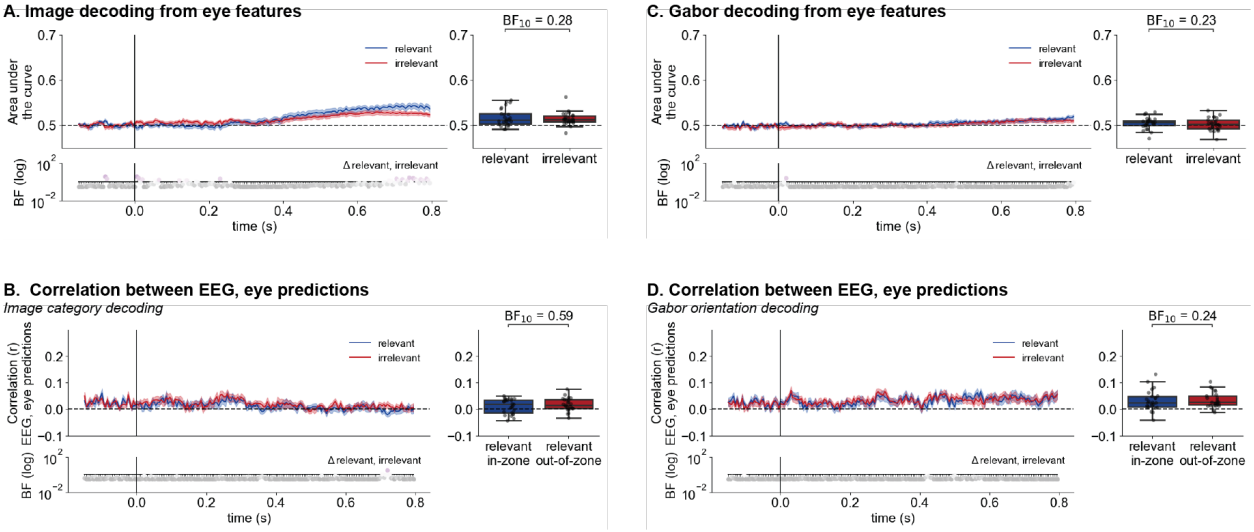



*Appendix 3 Figure 47. Confident and non-confident memory performance.* Memory performance was above-chance when using a less-stringent memory accuracy threshold, i.e., old stimuli reported as maybe old or definitely old were considered correctly-remembered.

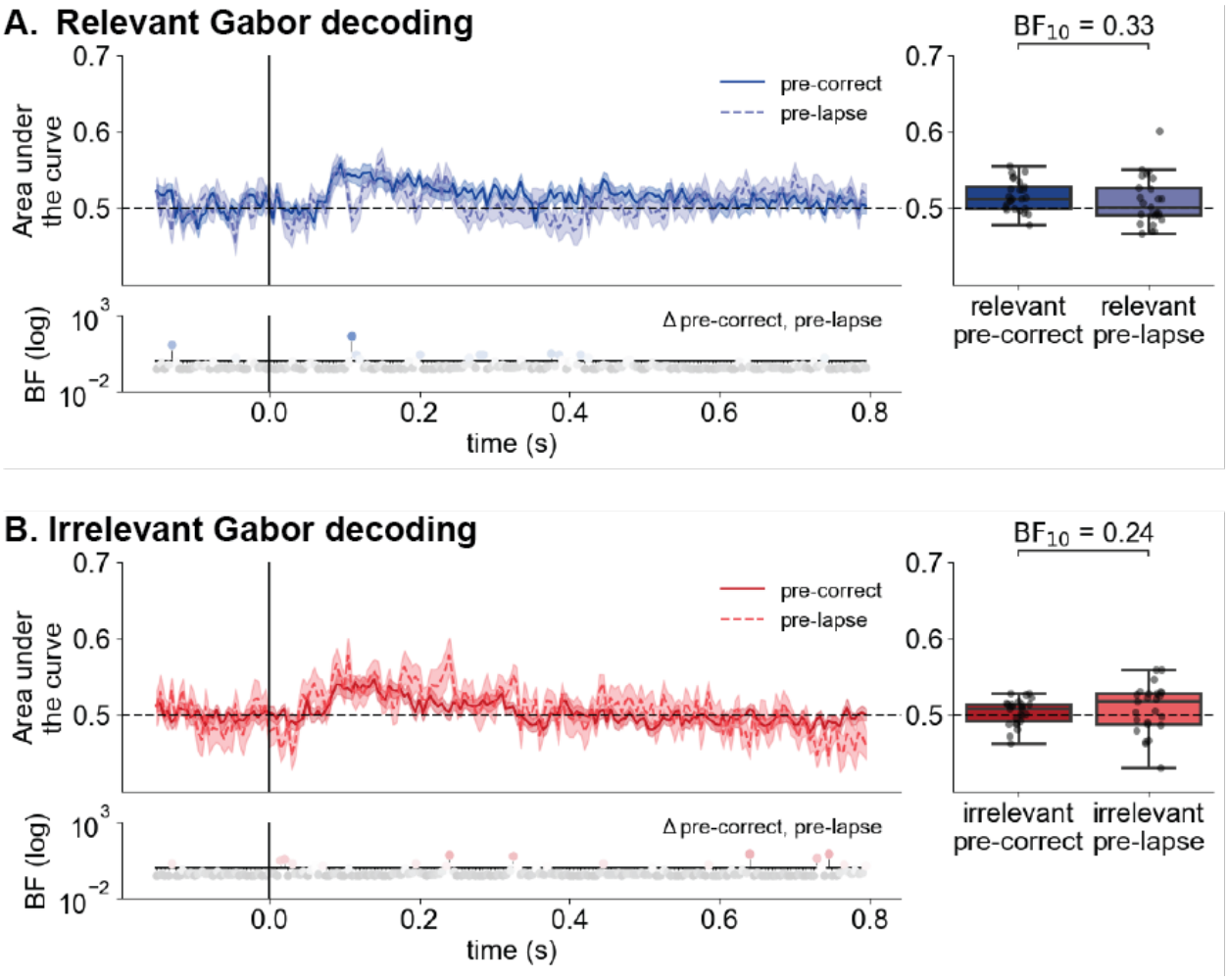
## Appendix 4: Chapter 6 supplementary materials



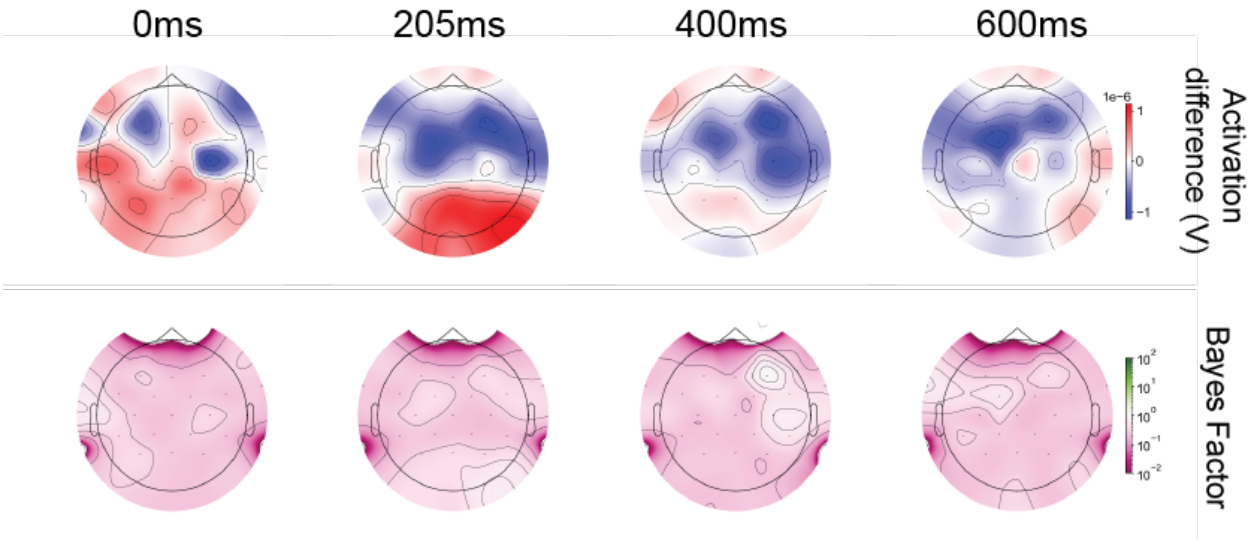

*Appendix 4 Figure 48. Relevant decoding does not exceed cross-validated decoding time courses.* We compared the time courses trained and tested during the localizer task, during which participants passively viewed stimuli, to those observed during the CPT. (A) When decoding image categories, relevant image decoding performance was very similar to the cross-validated localizer time course, whereas irrelevant image decoding was decreased. (B) When decoding Gabor orientations, the cross-validated localizer time course showed higher and more extended performance than during either Gabor-relevant or Gabor-irrelevant CPT blocks.



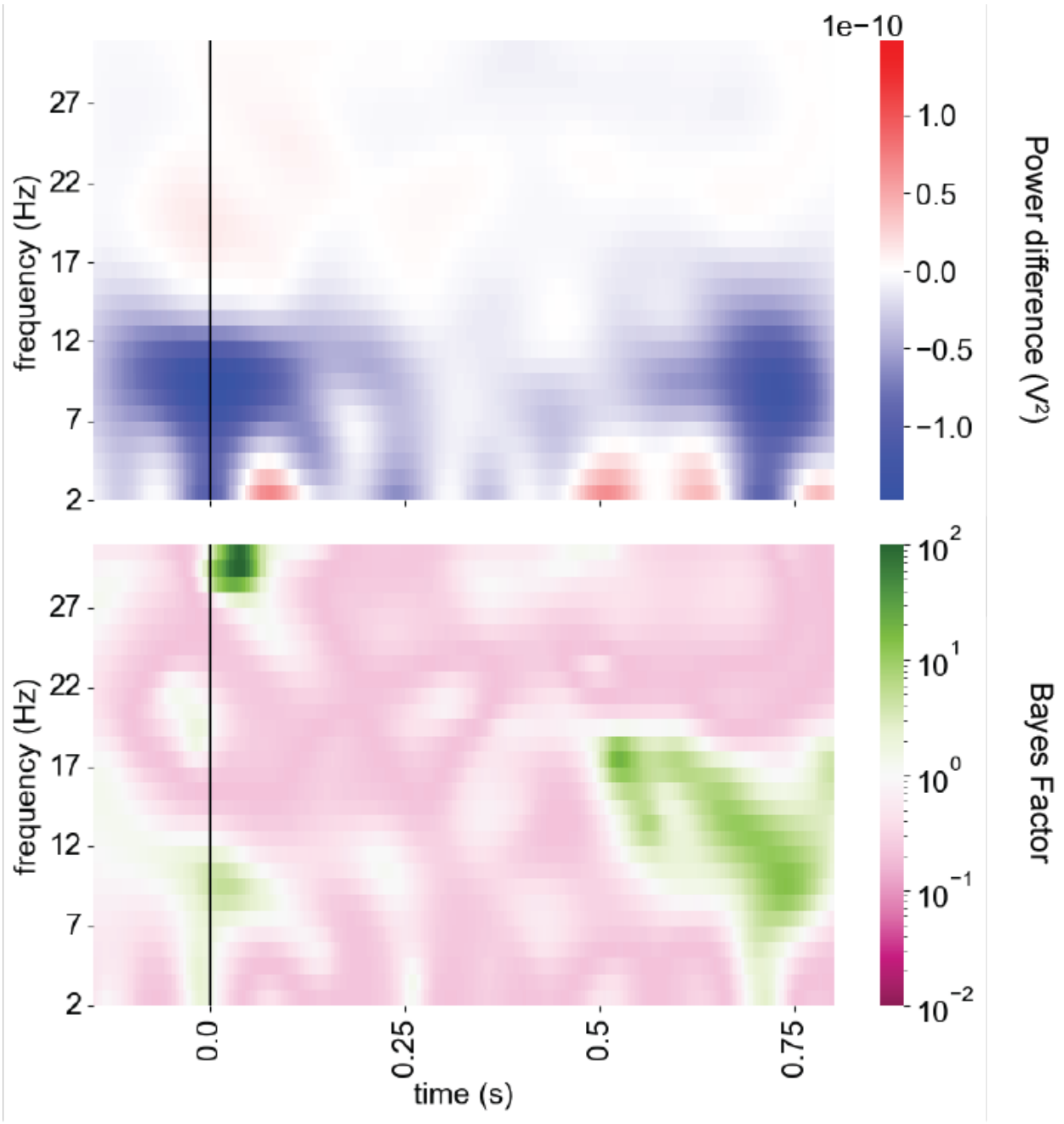

*Appendix 4 Figure 49. Decoding from eye features is distinct from EEG.* Classifiers using eye features resulted in distinct decoding time courses for (A) image category and (C) Gabor orientation relative to those trained on EEG signals. Further, predictions from EEG-based and eye-based classifiers were minimally correlated across trials for (B) images and (D) Gabors, and this correlation did not differ between task-relevant and task-irrelevant predictions. This suggests that differences in decoding accuracy observed as a function of attentional selection (task relevance) are not driven by differential dependence on eye-related features.




*Appendix 4 Figure 50. Gabor decoding does not differ preceding sustained attentional lapses.* Decoding performance for (A) relevant and (B) irrelevant Gabor orientations showed only small differences as a function of sustained attentional state. Trials under good sustained attention states were those preceding correctly withheld responses to infrequent target stimuli (pre-correct), while trials preceding lapses (pre-lapse) were considered to be in a poor sustained attentional state. We did not observe differences when averaging performance across the trial (0-800ms), as shown in the box-and-whisker plots.

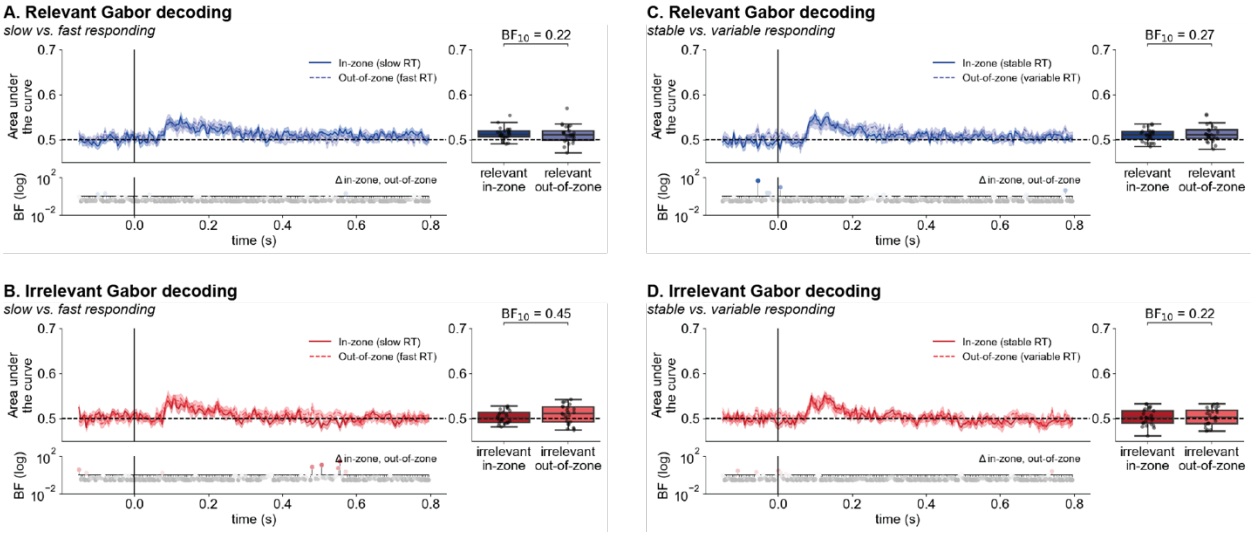



*Appendix 4 Figure 51. Minimal differences in event related potentials as a function of sustained attention.* Topographic maps reflect the difference in average activation across the scalp as a function of sustained attention (trials preceding correctly-withheld responses minus trials preceding lapses) in frequent images presented in the three trials preceding an infrequent target. Bayes factors reflect evidence of a reliable sustained attention-related activation difference at every channel. Bayes factors greater than 1 are plotted in green and reflect evidence of an effect.




*Appendix 4 Figure 52. Time-frequency decomposition for good vs. poor sustained attentional states.* Differences in time-frequency activity in posterior electrodes (good - poor sustained attention) reveals greater alpha (8-12 Hz) and theta (4-8 Hz) power at stimulus onset (time=0 s) and prior to stimulus offset (time=0.8 s) prior to sustained attentional lapses. Bayes factors reflect

evidence of an effect (distribution reliably different from 0) at every frequency and time point. Bayes factors greater than 1 are plotted in green and reflect evidence of an effect.




*Appendix 4 Figure 53. Minimal effect of response time-indexed sustained attentional state on Gabor orientation decoding.* We observed only small, transient differences in decoding performance for Gabor orientation as a function of response-indexed sustained attentional state using response time (RT) speed (A-B) and variance (C-D). When averaged across the trial, no differences were observed as a function of sustained attention. Box-and-whisker plots reflect subject-wise mean area under the curve across the trial (0-800 ms after onset).

| **Predictor** | **Mean estimated effect** | **89% Credible Interval** |
|---|---|---|
| Memorability | .131 | [.092, .170] |
| Paired lure false-alarm-ability | -.283 | [-.322, -.243] |
| Memory trial number | -.027 | [-.102, .047] |
| Relevance (relevant vs. irrelevant) | .231 | [.075, .386] |
| Frequency (infrequent vs. frequent) | .663 | [.550, .776] |
| Preceding RT speed | .025 | [.039, .143] |
| Preceding RT variance | -.020 | [-.022, .072] |
| RT speed:RT variance | -.020 | [-.040, $1.78*10^{-4}$] |

*Appendix 4 Table 14. Effects predicting both confident and non-confident memory judgements.* As a preregistered analysis, we tested what features predicted successful selection of the old stimulus during a two-alternative forced-choice memory task regardless of confidence rating. We fit a Bayesian mixed effects logistic model and report the mean estimated effect and 89% credible interval here. Credible intervals that do not include 0 are considered evidence of an effect. We observe that image memorability, the likelihood of the paired lure stimulus to elicit false alarms (paired lure false-alarmability), and stimulus frequency remain strong predictors of memory. Additionally, task-relevance and response speed emerge as predictors of memory that did not previously predict confident memory judgments alone.

| **Predictor** | **Mean estimated effect** | **89% Credible Interval** |
|---|---|---|
| Memorability | .017 | [$-2.53*10^{-3}$, .037] |
| Paired lure false-alarm-ability | -.089 | [-.109, -.069] |
| Memory trial number | $-1.65*10^{-4}$ | [-.020, .020] |
| Pre-trial EEG decoding performance | $-7.51*10^{-3}$ | [-.027, .012] |

*Appendix 4 Table 15. Effects predicting both confident and non-confident memory for infrequent target images.* We fit a Bayesian mixed effects logistic model predicting memory for rare target images where successful identification of the old stimulus is considered correct regardless of confidence. We report the mean estimated effect and 89% credible interval here. Credible intervals that do not include 0 are considered evidence of an effect. Only the likelihood of the paired lure image to elicit false alarms (paired lure false-alarm-ability) predicted image memory in this model.