THE UNIVERSITY OF CHICAGO

CHARACTERIZING WORKING MEMORY DELAY PERIOD ACTIVITY

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

NICOLE HAKIM

CHICAGO, ILLINOIS

JUNE 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF PUBLICATIONS

Hakim N, Simons DJ, Zhao H, & Wan X (2017) Do easterners and westerners differ in visual cognition? A preregistered examination of three visual cognition tasks. *Social Psychological and Personality Science, 8*(2), 142-152.

Hakim N, Vogel EK (2018) Phase-coded memories in mind. *PLOS Biology, 16*(8), e3000012.

Hakim N, Adam KCS, Gunseli E, Awh E, Vogel EK (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. **This article is reproduced as Chapter 2.**

Hakim N, Feldmann-Wüstefeld F, Awh E, Vogel EK (2020) Perturbing neural representations of working memory with task-irrelevant interruption. *Journal of Cognitive Neuroscience*, *32*(3), 558-569. **This article is reproduced as Chapter 3.**

Hakim N, deBettencourt MT, Awh E, Vogel EK (2020) Attention fluctuations impact ongoing maintenance of information in working memory. *Psychonomic Bulletin & Review, 27*(6), 1269-1278. **This article is reproduced as Chapter 1.**

Hakim N, Awh E, Vogel EK (2020) Manifold working memory. In Working memory: a multicomponent model (Eds. Robert Logie, Nelson Cowan, and Valerie Camos) *Oxford University Press*

Hakim N, Feldmann-Wüstefeld T, Awh E, Vogel EK (2021) Controlling the flow of distracting information in working memory. *Cerebral Cortex*. **This article is reproduced as Chapter 4.**

Hakim N, Awh E, Vogel EK, Rosenberg MD (submitted) Predicting cognitive abilities across individuals using sparse EEG connectivity. **This article is reproduced as Chapter 5.**

ACKNOWLEGEMENTS

I would like to acknowledge and thank those who played a role in my academic and scientific accomplishments. First of all, I would like to thank my advisors, Ed Vogel and Ed Awh, each of whom has provided patient advice and guidance. Throughout my PhD, I spent endless hours—in the lab and on zoom—talking to the Eds about science, new analyses, and more. Their dedication and keen interest in cognitive neuroscience have shaped my ideas and sharpened my scientific prowess. This dissertation would simply not have been possible without their guidance and expertise. I would also like to thank Monica Rosenberg for taking the time to work with me during her first years as a professor. Her enthusiasm and scientific rigor provided me with a new scientific perspective and refined my understanding of attention. Thank you also to Wilma Bainbridge for providing feedback and guidance on my dissertation. Additionally, thank you to all of my labmates and coauthors. They have undoubtedly improved my graduate experience by providing unwavering support and friendship. Finally, thank you to my friends and family—they should know that their support and encouragement is worth more than I can express on paper.

ABSTRACT

Temporarily holding information in mind is an important part of many cognitive processes. For example, when reading, you have to maintain multiple pieces of information in mind. Working memory is the mental workspace that is used to temporarily maintain and manipulate this kind of information. A central question in cognitive neuroscience is how this working memory system functions. In this dissertation, I use novel behavioral, neural, and computational methods to address this question and to expand our understanding of working memory theory. In Chapter 1, I provide evidence that maintaining information in working memory relies on multiple sub-processes, including storage capacity and attentional control. In Chapter 2, I characterize the multi-faceted nature of working memory by dissociating two distinct delay period signatures of working memory, the contralateral delay activity (CDA) and lateralized alpha power. In Chapters 3 and 4, I ask how these two delay period signatures of working memory respond to interruptions in order to further delineate their temporal dynamics. Finally, in Chapter 5, I build a model using delay period activity to predict individual differences in the amount of information someone can hold in working memory. Taken together, this research informs contemporary theories of working memory, supporting a working memory system that is comprised of multiple sub-component processes.

INTRODUCTION

The human brain is a powerful cognitive tool that has given rise to exceptional accomplishments: the development of life-saving vaccines, the establishment of civilizations, and the exploration of the moon. Our brains can store what seems to be a nearly infinite amount of information that we have learned over a lifetime, and we use this knowledge to construct a coherent sense of the world around us. Despite these exceptional capabilities, the human cognitive system is greatly restricted in terms of how much information can be actively held "in mind" at a given moment. That is, despite the general feeling that we have awareness of much of our immediate environment, we have access to only a tiny sliver of it—sampling the physical world around us in bite-sized amounts of information. This "working memory" system allows us to temporarily hold information in a readily accessible state so that we can utilize it to perform complex cognitive tasks, such as reasoning and language. It is often considered to be the "mental workspace" for thinking. It is also a primary source of variation across humans in overall cognitive abilities. For example, an individual's working memory capacity is a stable trait that is a strong predictor of their fluid intelligence, language skills, and scholastic achievement. One central question in cognitive neuroscience is how this working memory system functions. For example, what are the neural underpinnings of this system? Can we predict how much information one can hold in this "mental workspace" based on neural activity? In this dissertation, I *begin* to address some of these open questions.

Modern views of working memory explicitly characterize it within the context of its interactions with long term memory (Cowan, 1999). These embedded process models describe memory representations as existing in three potential states: inactivated long-term memory, including all representations stored in long term memory; activated long term memory, which are

latent representations that can quickly be brought into an active state due to contextual priming and recency; and the focus of attention, an active but sharply limited state in which only a small number of items can be represented simultaneously. These embedded processes models define working memory as any processing mechanism that allows information to be temporarily available. This functional definition of working memory highlights that an array of different mechanisms (e.g., the focus of attention, activated long-term memory, and inactivated long-term memory) can contribute to performance in working memory tasks. However, this definition of working memory makes it difficult to maintain the important theoretical distinction between working and long-term memory. How can this dissociation between working and long-term memory be maintained when many complex cognitive tasks rely on a careful collaboration between these two processes?

To gain leverage on this issue, we propose that active representations in working memory should be defined operationally based on neural activity. Previous research has shown that working memory generates persistent neural activity that tracks stimulus-specific information (Curtis & D'Esposito, 2003; Harrison & Tong, 2009; Vogel & Machizawa, 2004a). These processes are functionally distinct from long term memory because long term memories are thought to be maintained via changes in synaptic weights, not active neural processes (Lamprecht & LeDoux, 2004). Therefore, throughout this dissertation, I define working memory as active neural firing that supports the maintenance of information. Other types of memory that are supported by passive neural representations may be better understood as types of long-term memory. Thus, our neural definition of working memory most closely aligns with how embedded processes models define the focus of attention because both are active processes that

are sharply limited in capacity. In this dissertation, I will use the terms "working memory" and "focus of attention" synonymously.

Operational definitions of working memory, such as that which is endorsed by embedded processes models, essentially treats the active maintenance of information as a monolithic process. That is, information is simply either maintained or not. However, there is growing evidence that suggests that multiple distinct processes may contribute to successfully maintaining information in an active state. In Chapters 1 and 2, I will argue that working memory is manifold and should be thought of as a compilation of multiple processes that support the active representation of relevant information. Specifically, in Chapter 1, I build a computational model that dissociates the contribution of item-based storage and attentional control to working memory performance. In Chapter 2, I provide empirical evidence that working memory is comprised of an index of prioritized space and an item-based index. In Chapters 3 and 4, I further dissociate these distinct sub-component processes of working memory by interrupting them after information is already maintained in working memory. Finally, in Chapter 5, I build a predictive model based on neural activity during a working memory task to predict individual differences in working memory capacity.

Overall, there has been a longstanding effort to understand what exactly working memory is and how it is represented in the brain. In this dissertation, I offer a new hypothesis which proposes that working memory is multifaceted. I provide evidence for dissociable neural signals that separately track spatial and item-based information in working memory. Furthermore, I build a model that predicts individual differences in working memory capacity using neural and behavioral signatures measured during active maintenance. The evidence provided here further

elucidates how information is maintained in working memory, and my hope is that this will provide further traction for understanding this central component of intelligent human behavior.

PART I.

Distinct neural signatures of the working memory

CHAPTER 1.

Storage and attentional control distinctly contribute to working memory performance.

Visual working memory facilitates the temporary maintenance of information over time for use in ongoing cognitive processes. The majority of working memory research has examined performance over short (~1 second) retention intervals. However, there is growing evidence that suggests that working memory performance declines after longer (~10 seconds) retention intervals (Donkin et al., 2015; Rademaker et al., 2018; Zhang & Luck, 2009a). These studies have mostly focused on whether this decline in performance is due to changes in precision or due to the sudden loss of information over time. In the present work, however, we will focus on an orthogonal distinction that further characterizes how increasing retention intervals affect performance. Specifically, we will dissociate how attentional control and working memory capacity contribute to working memory performance over long retention intervals.

Previous work has investigated the intricate relationship between attention and working memory performance (Adam et al., 2015a; Awh et al., 2006; De Fockert et al., 2001; deBettencourt et al., 2019). This work has provided evidence that trial-by-trial variability in attention contributes to individual differences in working memory performance. One such study deployed a whole report procedure, in which participants reported the identity of each item from the memory display on every trial (Adam et al., 2015a). They found that differences in visual working memory performance both across and within individuals over short retention intervals were well described by a combination of two separate parameters: *Kmax,* maximum working memory capacity, and *a*, a factor related to attentional control (Adam et al., 2015a). Given that

these two parameters, *Kmax* and *a*, explain performance variability at short retention intervals, in our study, we are interested in whether the decline in performance observed after long retention intervals is due to changes in *Kmax*, *a*, or some combination of these two parameters.

Another unresolved concern is that we do not know how long retention intervals act on multiple representations because previous research has utilized partial report tasks, in which a single item from the display is randomly chosen to be tested. In particular, there are two possible alternatives for performance declines at long retention intervals: (1) participants could drop all information from working memory (*complete drop*) or (2) participants could drop individual items (*partial drop*). Longer retention intervals could increase the probability that participants completely disengage from the task, causing participants to lose all information on a subset of trials. Previous research has found that information is lost suddenly and completely when participants have to maintain information for extended amounts of time (Zhang & Luck, 2009b). This research investigated performance by probing one item at a time, so they could not draw conclusions about the maintenance of all items on a single trial. Nevertheless, their findings suggest that participants could suddenly and completely lose all information from working memory when they have to maintain information over longer (10s) amounts of time. This hypothesis aligns with the complete drop model. Alternately, the efficiency of attentional control could decrease over time. This could cause a partial loss of information over long retention intervals. One task that is well suited to examine representations of multiple items on a single trial is a discrete whole report working memory task. This task requires participants to click among discrete colors to report the identity of each item in the display (Adam et al., 2015a). Performance on discrete whole report tasks can range from getting no items correct, to getting all items correct, or anything in between. Across individuals, average discrete whole report

performance is highly correlated with performance on traditional change detection tasks, where a single item is randomly probed (Adam et al., 2015a). However, the discrete whole report task has substantial advantages because it measures memory for each item on a multi-item display. This allows us to assess whether changes in performance after long retention intervals are due to completely dropping all information from working memory or due to an increased likelihood of dropping individual items.

In the current study, we had participants perform a discrete whole report task while maintaining information over long (10 s) and short (1.5 s) retention intervals. To analyze the data, we designed a computational model that simultaneously solved for attentional control (*a*) and working memory capacity (*Kmax*) using grid search. By simultaneously fitting both of these parameters, we improved upon and extended previous research (Adam et al., 2015a). We were interested in (1) characterizing how the distributions in the number of items remembered varied across short vs. long retention intervals, and (2) using these distributions to model how attention and working memory capacity were impacted by the length of the retention interval.

Experiment 1-1

In Experiment 1-1, we characterized trial-by-trial fluctuations in the number of items maintained over time, while distinguishing between effects on the maximum number of items that could be stored, and an attention parameter that quantifies the probability that participants would achieve that maximum. Participants performed a discrete whole-report task which probed memory for each item on every trial, following either short (1.5 s) or long (10 s) retention intervals. We had participants maintain information over short (1.5 s) and long retention intervals (10 s) in order to determine whether longer retention intervals lead to declines in maximum

working memory capacity (*Kmax*), or the probability of achieving *Kmax* (*a*), which we will refer to as attentional control. Previous research has numerically solved for *Kmax* and *a* by running simulations (Adam et al., 2015a). Here, we solved for these two parameters using grid search. In addition, we examined whether failures to achieve maximum capacity were best characterized by a complete drop of all information from working memory, or a partial drop of some items while maintaining other items.

Methods

Participants. Twenty adults participated in Experiment 1-1 (8 male, 19–31 years, mean 23.25 years). Two additional participants were excluded for failing to complete the experiment in the allotted time (2 hours), resulting in an insufficient number of trials per condition (<180). The final sample size for both experiments was selected a priori based on previous discrete whole-report sample sizes (Adam et al., 2017). All participants in both experiments received payment (US $20, $10/hour) or course credit (2 credits, 1 credit/hour). Additionally, all participants reported normal or corrected-to-normal visual acuity and color vision and provided informed consent to a protocol approved by the University of Chicago Institutional Review Board.

Stimuli. Participants encoded arrays of six differently colored squares (subtending approximately 1°×1°) on a gray background. Each square was one of nine possible colors without replacement (red, green, blue, yellow, magenta, cyan, white, black, orange). Therefore, each memory array consisted of six squares of six different colors. For the response screen, multicolored squares appeared that were a 3×3 grid of all nine colors. Participants were instructed to fixate on a black central dot (0.43°).

Apparatus. Stimuli were generated in MATLAB (The MathWorks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997). Participants were seated approximately 67 cm from

a 24-inch LCD BenQ monitor. Squares could appear within approximately a 20°×20° area on the screen. To monitor compliance with the articulatory suppression instructions, audio was recorded through MATLAB with an Eberry microphone.

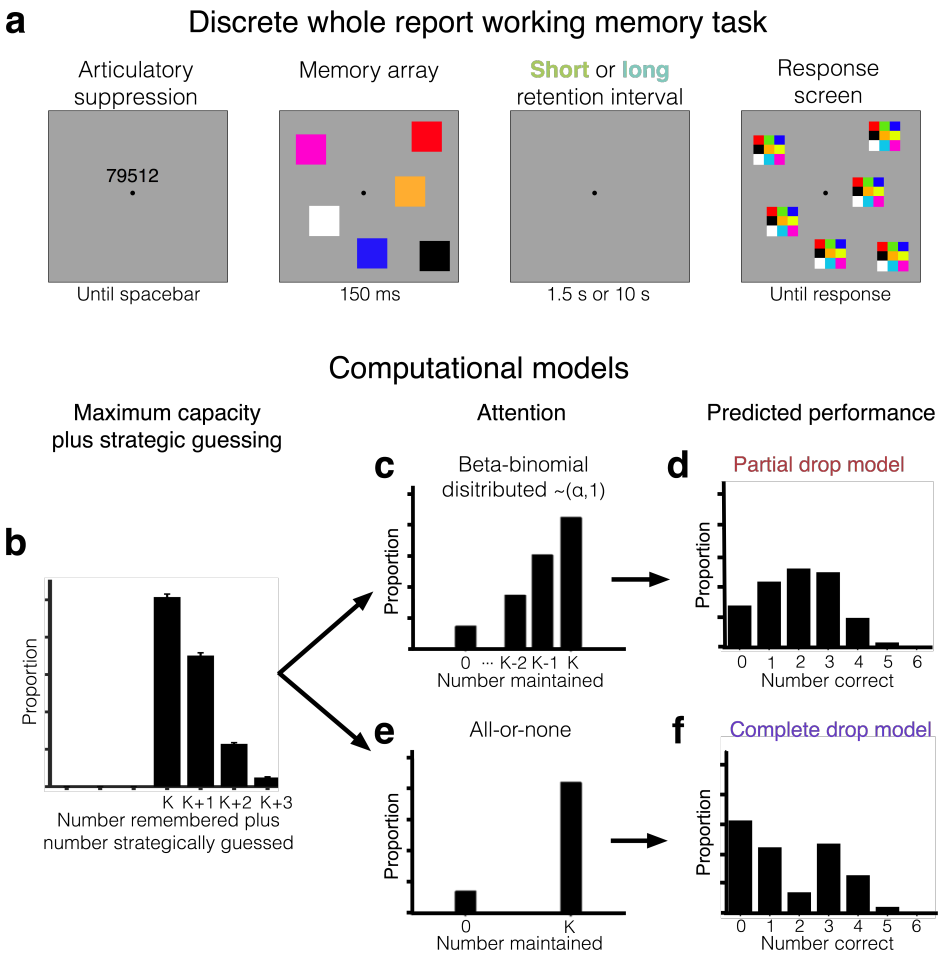

Figure 1-1. Task design and computational model.

A) Discrete whole-report working memory task for Experiment 1-1. At the start of each trial, five digits appeared on the screen. Participants said these numbers at least once before trial initiation and repeated them continuously until the response screen appeared. Participants initiated each trial by pressing spacebar, and then a memory array of six colored squares appeared briefly (150 ms). The critical manipulation was the length of the retention interval, either short (1.5 s) or long (10 s). At the end of the retention interval, a response screen appeared with multicolored squares at the location of each original square. Participants selected the color of each square using the mouse. Participants were required to respond to all six items before the trial would proceed, and the multicolored squares remained on the screen until the participant finished responding. B) Both models assume that participants have a stable maximal capacity, plus guessing. These models assume that participants strategically guessed the colors of the items that they did not maintain in memory. If a participant remembers three items, for example, then

(Figure 1-1 continued) they are able to narrow the possible colors of the remaining three items from 1/9 to 1/6 possible colors. If participants correctly guessed the colors of the items that they did not maintain in memory, then working memory performance on that trial could exceed working memory capacity (*Kmax*). C) In the partial drop model, attentional control is distributed according to a beta-binomial function. D) The partial drop model is unimodal. It reflects a partial loss of information, as participants can drop individual items, as opposed to the whole array. E) In the complete drop model, attentional control is distributed according to a Bernoulli function. Attention to the array is therefore essentially all-or-none, as participants are either fully engaged and remember *Kmax* items, or they are fully disengaged and do not remember any of the items. F) In the complete drop model, performance typically reflects a bimodal distribution. The distribution becomes unimodal if participants have extremely low working memory capacities.

Procedure. Participants completed a discrete whole-report visual working memory task with articulatory suppression (Figure 1-1). Prior to the onset of the memory array, five randomly generated digits appeared above the fixation dot. Participants said the numbers out loud at least once before pressing spacebar to begin the trial and continued to repeat the numbers until the response screen appeared. Participants encoded memory arrays of six colored squares that appeared briefly (150 ms), and then maintained these squares over a blank retention interval. Critically, the retention interval was either short (1.5 s, 50% of trials) or long (10 s, 50% of trials). After the retention interval, a response screen appeared with multicolored squares at the location of each original square. To respond, participants selected the color of the original square at each location using the mouse. Participants were required to respond to all 6 items before the trial would proceed, and all squares remained on the screen until the participant finished responding. Participants completed a block of 60 trials and then took a 30 s break. Within each block, the retention interval length was held constant (short or long). In total, participants completed 360 trials.

Analysis. Working memory performance was quantified per trial as the number of items for which the correct color was selected, ranging from 0 (no items correct) to 6 (all items correct). Across trials, we calculated the mean number of correct responses.

Computational models. In order to describe performance, we applied a family of computational models that have two parameters, one for attentional control (*a*) and one for maximum working memory capacity (*Kmax*) (Figure 1-1, b-f). These models assume that performance can be described using (1) a stable maximal capacity (2) attentional control and (3) strategic guessing.

In these models, each participant has a stable maximal capacity. In line with work that suggests that participants remember entire objects (L), capacity is an integer value that can vary from one item to the maximum number of items in the display, with a step size of one ($Kmax \in [1,2,3,4,5,6]$). On any given trial (i), participants maintained some integer number of items (N_i) in memory, ranging from 0 to their maximal capacity ($N_i \in [0,1,2,\ldots,Kmax]$). To investigate whether an integer value was critical for these results, we re-ran the model using a smaller step size of ($\Delta Kmax = 0.5$) and found the same pattern of results.

We further assumed participants would strategically guess for the remaining items not maintained in memory. That is, for a given trial (i), participants maintained an integer number of items with the correct color in mind (N_i). For the remaining items ($6–N_i$), we used a binomial distribution to model strategic guessing from among the remaining colors ($9–N_i$).
These models also include a term for attentional control throughout the delay. Attentional control was operationalized as the probability of maintaining the maximum number of items (*Kmax*) for each trial.

In particular, this model distinguishes between working memory performance and working memory capacity. Maximal working memory capacity (*Kmax*) is the maximum amount of information that an individual can maintain in working memory at any one moment. Note that this operationalization of maximal working memory capacity is related to, but distinct from,

traditional measures of average performance, i.e., K (Cowan, 2010; Luck & Vogel, 2013). On each trial, the number of correctly reported items can range from 0 to 6 items. Participants could correctly report more items than *Kmax* items if they correctly guess the color of any of the items that they did not maintain in working memory. They can report fewer items than their *Kmax* if they have low attention control (*a*) on that trial. We examined two conceptualizations of attentional control in the different models:

In the *complete drop model*, attentional control was effectively binary: participants either maintained their maximum number of items (N_i=*Kmax*), or they had no information at all (N_i=0). This model effectively reduces to a Bernoulli distribution over values zero and *Kmax* with the probability of *Kmax* being one minus the probability of potential failure.

In the *partial drop model*, the number of maintained items ranges over the integer numbers from 0 to *Kmax*. In this situation in statistics, we would typically apply the binomial distribution, which gives the probability distribution of zero to *Kmax* successes given a constant probability of success on each trial. However, we believe that the probability of success throughout an experiment is not constant. Prior evidence has shown that attentional control fluctuates over the course of an experiment and these attention fluctuations lead to fluctuations in working memory performance (deBettencourt et al., 2019). Therefore, we chose to use the beta-binomial distribution because this allowed the probability of success to also follow a probability distribution, rather than remain constant on each trial. For *n* trials, this beta-binomial distribution is characterized by a beta function, $B(\alpha, \beta)$, and a binomial coefficient $\left(\frac{n}{Kmax}\right)$, with the following probability density function:

$$P(Kmax|n, \alpha, \beta) = \frac{B(\,Kmax + \alpha, n - Kmax + \beta)\left(\frac{n}{Kmax}\right)}{B(\alpha, \beta)}$$

Model fitting. We computed the probability distributions for the complete and partial drop models over the 2D grid space of the free parameters for each model: *Kmax*∈[1,6] (with a step size of 1) and a∈[0,10] (with a step size of 0.01). The complete drop model had one free parameter, p, which represented the probability of disengagement. The partial drop model also had one free parameter, α, which represents the robustness of attentional control. Alpha determines the mean of the beta distribution for p by $mean = \frac{\alpha}{\alpha+\beta}$. For this model, we kept the β parameter equal to 1, so that the beta distribution was [α, 1] for each trial. We were interested in the mean of the distribution, so we fixed β equal to 1 and kept α free. By fixing β to 1, we assume that attentional control fluctuates quite a bit, especially when the success probability is intermediate. We then determined the best-fitting model to each individual participant's performance distributions by calculating which combination of parameters produced the largest log likelihood (*LL*). We calculated *LL* separately for each participant and model. We then sum the *LL* for all participants for a given model (e.g. $LL_{complete}$). To compare the complete and partial drop models, we used the Bayesian Information Criterion (BIC). We calculated the difference in BIC (ΔBIC) values between the two models. In this calculation, the penalty term of the BIC drops out because both models have the same number of free parameters (i.e. 1). We used the following formula to calculate the difference in BIC values between the two models:

$$\Delta BIC = -2 \times (LL_{complete} - LL_{partial})$$

$LL_{complete}$ is the log-likelihood of the complete drop model, and $LL_{partial}$ is the log-likelihood of the partial drop model. Positive values of ΔBIC indicate a better fit for the partial than the complete drop model.

To further investigate performance, we fit two combined partial drop models that aggregated across data from short and long retention intervals with three total parameters. Here,

we manipulated which parameter (*a* or *Kmax*) was fixed or free to vary across long and short retention intervals. In the first model ("*a*-free model"), attentional control (*a*) was free to vary across retention intervals, while maximum working memory capacity (*Kmax*) was fixed. In the second model ("*Kmax*-free model"), maximum working memory capacity (*Kmax*) was free to vary across retention intervals, while attentional control (*a*) was fixed. We used the following formula to calculate the difference in BIC values between these two models:

$$\Delta BIC = -2 \times (LL_{Kmax-free} - LL_{a-free})$$

$LL_{Kmax-free}$ is the log-likelihood of the *Kmax*-free model, and LL_{a-free} is the log-likelihood of the *a*-free model. Positive values of ΔBIC indicate a better fit for the *a*-free than the *Kmax*-free model.

Statistics. Results are reported as the mean plus or minus the standard deviation, unless noted otherwise. Statistics in all experiments were computed using Student's two-tailed paired *t*-tests. Effect sizes are reported as Cohen's *d*.

Results

Behavioral performance.

Working memory performance (Figure 1-2) was operationalized as the number of items correct per trial. Average working memory performance (*n*) was calculated for each individual across all trials. If participants had been randomly guessing, chance performance would equal the probability of getting any item correct (1 of 9) multiplied by the number of items (6, chance=0.67). Working memory performance across both conditions was well above chance (n=2.40±0.85; $t(19)$=10.00, p<0.001).

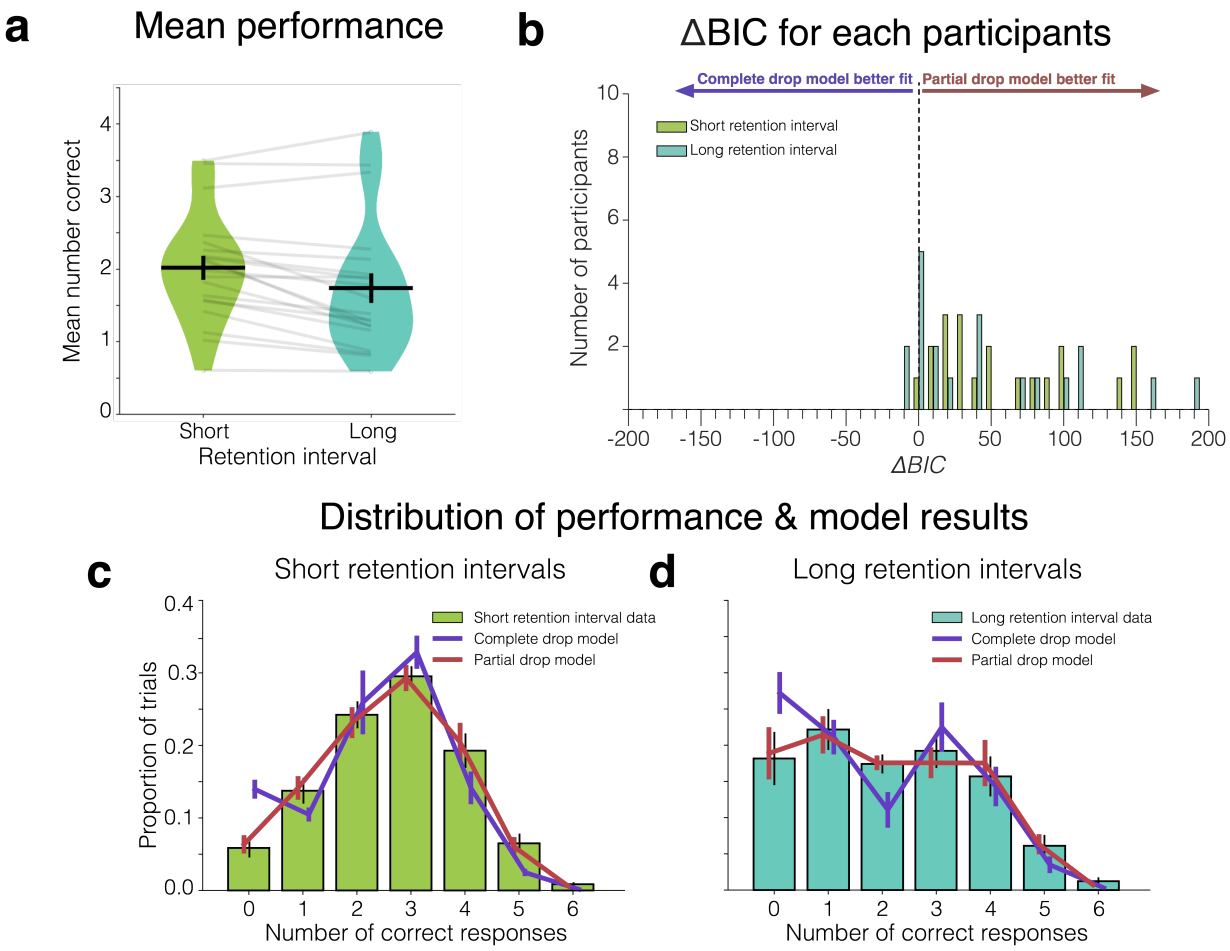

Figure 1-2. Behavioral results and computational model fits for Experiment 1-1.

A) Working memory performance for short (green) vs. long (blue) retention intervals. Each dot represents the mean number correct in one condition for one participant. A line connects data from the same participant. The horizontal black line depicts the mean across participants. The black error bars reflect the standard error of the mean. The shaded area reflects the density across participants. B) Histogram of the difference between the BIC values (Δ*BIC)* in the complete and partial drop models for each participant. Positive values indicate a better fit for the partial drop model. The bars are centered within each bin (bin size =10) for the short (green) and long (blue) retention intervals. C) The green bars of the histogram depict the data from short retention interval trials. The height of each bar is the proportion of trials for each number of correct responses. Black error bars are the standard error of the mean across participants. The lines reflect the model fits using the best fitting parameters: complete drop model (purple) and partial drop model (red). Each model was fit independently for each participant, and error bars reflect the standard error of the mean of the best fitting parameters across participants D) The data (blue bars) and best fitting models (purple and red lines) from long retention interval trials.

The critical question was how performance differed over time (long vs. short retention intervals; Figure 1-2, a). In line with prior work, we hypothesized that performance would decline following long retention intervals (Zhang & Luck, 2009a). We calculated the average number of correctly remembered items separately for trials with long and short retention intervals (n_{long}=2.15±1.00, n_{short}=2.66±0.59). Indeed, participants remembered significantly fewer items after long vs. short retention intervals ($t(19)=4.35$, $p<0.001$, Cohen's $d=0.62$).

Computational modeling.

We developed an analytical solution to two computational models that could describe performance distributions: *complete* and *partial drop models*. These models make distinct predictions about the distribution of the number of correct responses across trials. The complete drop model was "all-or-none," such that participants either performed at their maximum capacity and remembered *Kmax* items, or they had no information about any of the items in the array, and, thus, maintained zero items. The partial drop model, on the other hand, assumed graded performance, such that participants could maintain anywhere from zero to *Kmax* items in working memory on any given trial. When *Kmax* is equal to 0 or 1, these models make similar predictions.

First, we determined whether the complete or partial drop model better explained our data. To do this, we compared the fits between the models using a goodness-of-fit measure, *BIC*. We took the difference between BIC values for the partial drop and the complete drop models, so positive Δ*BIC* values indicate that the data are better fit by the partial drop model, whereas negative Δ*BIC* values indicate better fit by the complete drop model. In this experiment, both models made reasonable fits to data (Figure 1-2, b-c). However, data from both short and long

retention intervals (Figure 1-2, b-c) were better fit by the partial drop model (ΔBIC_{long}=1042.86, ΔBIC_{short}=1298.73) than the complete drop model. For the short retention intervals, 19 out of 20 participants were better fit by the partial drop model and 1 out of 20 participants was better fit by the complete drop model. For long retention intervals, 15 out of 20 participants were better fit by the partial drop model, 3 out of 20 participants were better fit by the complete drop model, and 2 out of 20 participants were equally well fit by both models. This suggests that performance fluctuates along a continuum and these fluctuations affect the maintenance of individual items in an array.

Finally, the central question of interest was whether poorer performance following long retention intervals was best explained by a reduction in maximum working memory capacity or a reduction in the probability of achieving that maximum. The partial drop model showed that attentional control was reliably lower for long relative to short retention intervals (a_{long}=1.28±2.20, a_{short}=2.44±2.84; $t(19)$=3.30, p=0.004; Cohen's d=0.74). However, there was not a reliable difference in maximum capacity ($Kmax_{long}$=3.55±0.55, $Kmax_{short}$=3.20±0.36; $t(19)$=–1.92, p=0.07; Cohen's d=–0.43). These results suggest that the performance decrements after long retention intervals reflect an inability to maintain attentional control throughout the longer retention intervals, not a reduction in the maximum number of items that could be stored in working memory.

To more formally investigate how performance varied across long and short retention intervals, we fit two combined models. We alternated which parameter (*a* or *Kmax*) was free to vary across retention interval length. In the first model (*a*-free model), we allowed attentional control to vary across retention intervals while maximum working memory capacity was a fixed value for each individual (*a*-free, *Kmax*-fixed). In the second model (*Kmax*-free model), we

allowed maximum working memory capacity to vary across retention intervals while attentional control was a fixed value for each individual (*a*-fixed, *Kmax*-free). We observed that the a-free model better fit the data (ΔBIC=17.15). This provides further evidence for the finding that attentional control (*a*) varied across short and long retention intervals, whereas maximum working memory capacity (*Kmax*) was stable.

Conclusions

We developed a task that manipulated the length of the retention interval in a discrete whole-report working memory paradigm. Average performance declined after long vs. short retention intervals. We developed two models: the complete and the partial drop models, that both describe the number of items remembered as a result of some combination of attentional control and maximum working memory capacity. We found that the partial drop model better accounted for performance, which suggests that the amount of information in working memory across trials varied continuously. Additionally, the attentional control parameter reliably differed across retention intervals, such that it was lower after long retention intervals. However, the maximum working memory capacity parameter was equivalent for long and short retention intervals. Thus, decrements in performance after long retention intervals are due to a graded decline in sustaining attentional control, which leads to loss of individual items from working memory.

Experiment 1-2

In Experiment 1-2, we sought to replicate Experiment 1-1 and extend our understanding of the role of attentional control during working memory maintenance. We interpreted the findings from Experiment 1-1 as reflecting variations in sustained attentional control during

retention of information in working memory. In Experiment 1-1, however, we manipulated retention interval length across blocks, so, the differences that we found in Experiment 1-1 could be due to differences in preparation. For example, if participants know that they are going to have to retain information over a very long retention interval, they could better prepare than if they did not know the duration of the upcoming retention interval. To further eliminate potential differences in preparation across blocks, we intermixed trials of short and long retention intervals within blocks.

Methods

Participants. Twenty adults participated in Experiment 1-2 (10 male, 18–33 years, mean 24.05 years). Four additional participants were eliminated for technical issues with the experimental display code and four additional participants were excluded from analyses for leaving the experiment early.

Stimuli and apparatus. Same as Experiment 1-1.

Procedure. Same as Experiment 1-1, except short and long retention intervals were randomly inter-mixed across trials within each block. Participants completed a block of 48 trials and then took a 30 s break. In total, participants completed 288 trials.

Analysis, computational modeling, and statistics. Same as Experiment 1-1.

Results

Behavioral performance.

Just as in Experiment 1-1, working memory performance (Figure 1-3) was operationalized as the number of items correct per trial, and average performance (*n*) was calculated for each individual across all trials. Working memory performance on the discrete

whole-report task was significantly above chance (n=1.88±0.83; $t(19)$=6.61, p<0.001) suggesting that participants were not randomly guessing.

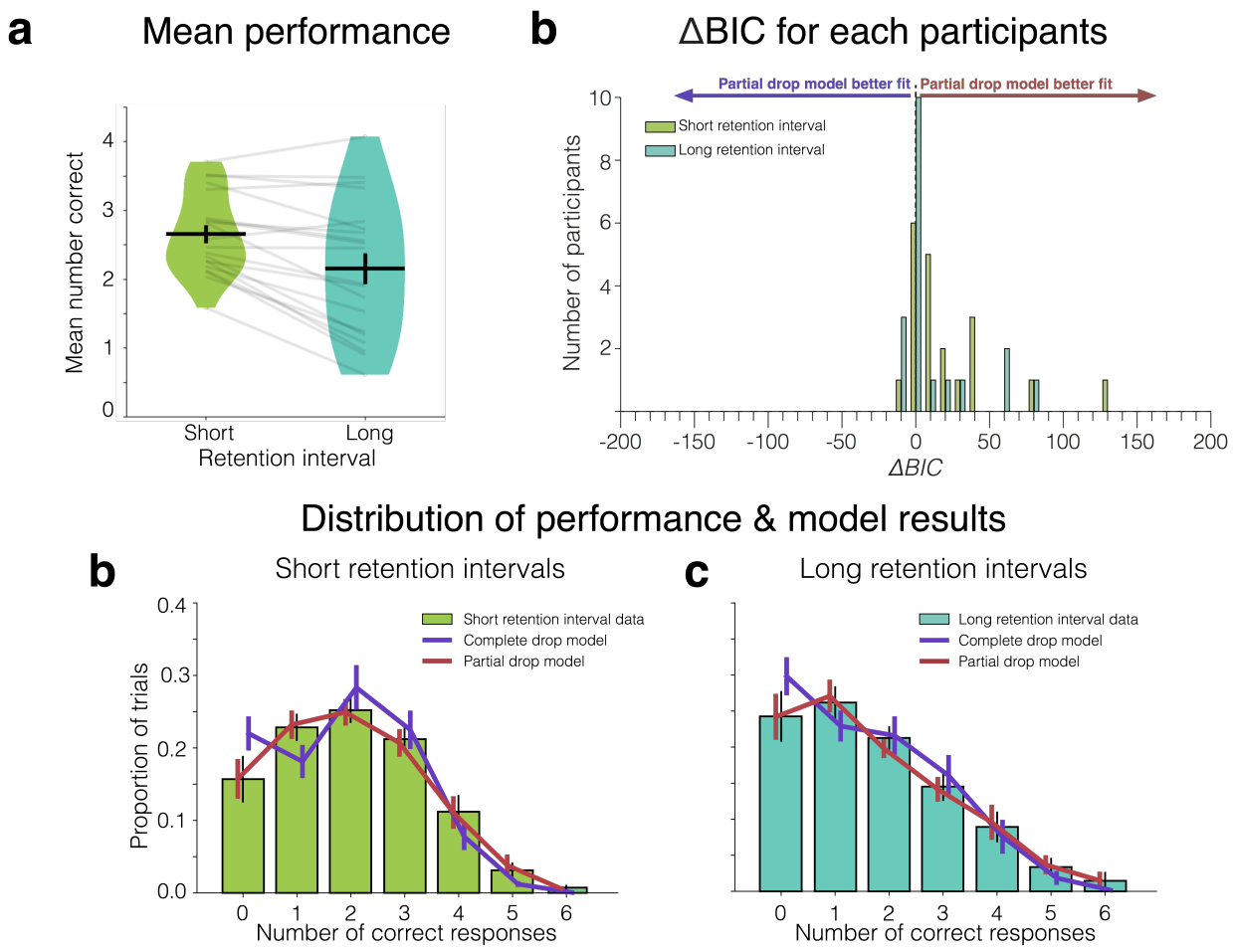

Figure 1-3. Behavioral results and computational modeling fits for Experiment 1-2.

A) Working memory performance for short (green) vs. long (blue) retention intervals. Each dot represents the mean number correct in one condition for one participant. A line connects data from the same participant. The horizontal black line depicts the mean across participants. The black error bars reflect the standard error of the mean. The shaded area reflects the density across participants. B) Histogram of the difference between the BIC values (*ΔBIC)* in the complete and partial drop models for each participant. Positive values indicate a better fit for the partial drop model. The bars are centered within each bin (bin size =10) for the short (green) and long (blue) retention intervals. C) The green bars of the histogram depict the data from short retention interval trials. The height of each bar is the proportion of trials for each number of correct responses. Black error bars are the standard error of the mean across participants. The lines reflect the model fits using the best fitting parameters: complete drop model (purple) and partial drop model (red). Each model was fit independently for each participant, and error bars reflect the standard error of the mean of the best fitting parameters across participants.

(Figure 1-3 continued). D) Data (blue bars) and best fitting models (purple and red lines) from long retention interval trials.

The critical question was whether working memory performance would again decline when the retention intervals were longer and whether this decline was due to changes in the attentional control parameter or due to changes in maximum capacity (Figure 1-3, a). We again observed worse performance after long retention intervals (n_{long}=1.74±0.91, n_{short}=2.02±0.75; $t(19)$=3.73, p<0.001, Cohen's d=0.34).

Computational modeling.

We first compared our two computational models to determine whether the complete or partial drop model better explained our data. Once again, we found that both models made reasonable fits to the data (Figure 2, b-c). However, just as in Experiment 1-1, we found that data from both the short and long retention intervals were better fit by the partial drop model (ΔBIC_{long}=590.38, ΔBIC_{short}=542.92) than the complete drop model. This finding reaffirms that performance across trials varies along a continuum and participants drop certain items from working memory while maintaining others over time. In fact, for the short retention interval trials, all participants were better fit by the partial drop model than the complete drop model. For the long retention interval trials, 17 out of 20 participants were better fit by the partial drop model, 2 out of 20 were better fit by the complete drop model, and 1 out of 20 participants were equally well fit by both models.

Next, we investigated the parameter fits for attentional control and maximum working memory capacity for the partial drop model. We again found that attentional control was reliably lower during long than short retention intervals (a_{long}=0.73±0.56, a_{short}=1.15±0.47; $t(19)$=4.59,

$p<0.001$, Cohen's $d=1.03$). Additionally, there was not a reliable difference across retention intervals between maximum capacity ($Kmax_{long}=3.00\pm1.30$, $Kmax_{short}=2.95\pm0.65$; $t(19)= -0.24$, $p=0.82$; Cohen's $d=-0.05$).

Just as we did in Experiment 1-1, we additionally fit two combined models. One of these models (*Kmax*-free model) varied *Kmax* and kept *a* constant, while the other model (*a*-free model) varied *a* and kept *Kmax* constant. These models allowed us to account for behavior across short and long retention intervals at the same time. We found that the *a*-free model fit the data better than the *Kmax*-free model ($\Delta BIC=4.31$). The comparison of these two models, once again, confirmed that maximum working memory capacity is stable across short and long retention intervals, whereas attentional control varies.

Conclusions

In Experiment 1-2, we eliminated potential differences in preparation by randomly intermixing trials and, therefore, preventing participants from preparing for a certain length retention interval ahead of time. We replicated all of our findings from Experiment 1-1. Performance declined after long (vs. short) retention intervals. This decline in performance was graded, such that the amount of information in working memory across trials varied continuously. The maximum amount of information that could be maintained in working memory remained the same over both long and short retention intervals. However, the probability of maintaining this maximum amount of information in working memory declined for long compared to short retention intervals.

Discussion

Past work has shown that variability in attention and filtering at encoding is an important limiting factor for working memory performance (Adam et al., 2015a; Fukuda & Vogel, 2011; Vogel et al., 2005). That said, the decline in working memory performance with longer retention intervals – for which encoding demands are perfectly matched – demonstrates that information is also lost even after successful encoding into working memory. Here, our main goals were to (1) characterize how the distributions of the number of items remembered varied across long and short retention intervals, and (2) use these distributions to characterize how the length of the retention interval impacts forgetting over time. Firstly, over two experiments, we observed that performance on a discrete whole-report task declined as retention interval increased (1.5 versus 10 seconds). Secondly, we found evidence in favor of the partial drop model, in which information loss over time is best described as a probabilistic loss of individual items rather than a complete failure of storage during specific trials. We developed a refined computational model of working memory performance that simultaneously solves for attentional control (*a*) and a maximum capacity (*Kmax*) by using grid search. We observed that when participants had to maintain information for a longer amount of time, the maximum amount of information (*Kmax*) that they could store in working memory remained the same. However, the probability of sustaining attentional control (*a*) was lower after long compared to short retention intervals.

In both Experiment 1-1 and 1-2, an individual's maximum working memory capacity (*Kmax*) was stable across short and long retention intervals. These results advocate for a perspective where maximal working memory capacity is relatively invariant across the population. Furthermore, there was relatively little range in maximal working memory capacity, suggesting that individuals differed not in the total maximum amount of information that they stored, but in

the frequency with which they stored that information. An extension from these findings is that maximum working memory capacity might be an immutable variable, with little range across the population. However, one caveat to this interpretation is that we observed numerically lower values of maximal working memory capacity in Experiment 1-2, when retention interval length was inter-mixed. Therefore, there might be other factors, such as experimental design, that influence attentional control and/or maximum working memory capacity. Future work could address when and how maximum working memory capacity (*Kmax*) changes.

We found that performance decrements after long retention intervals were not driven by reductions in maximum capacity. However, precision of representations in working memory could still change over time, as has been previously suggested (Fougnie, 2008; Rademaker et al., 2018; Schneegans & Bays, 2018). Our task provides a coarse measure of individual working memory representations and, thus, we cannot draw strong conclusions about changes in precision. However, even if precision does change over time, it may not lead to an aggregate increase in forgetting rates at the level of individual items. Future work could investigate how working memory precision changes over time by using a whole report task where participants report the precise color of every representation on every trial (Adam et al., 2017).

In sum, with this series of experiments, we extend our understanding of the impact of retention interval length on working memory performance. Past work has tended to portray capacity limits in working memory as reflecting a "ceiling" on storage capacity, similar to the limited space inside a container. From that perspective, it is plausible that factors that impair working memory performance yield a reduction in the available space in working memory. By contrast, we highlight recent work that suggests that variations in working memory performance across individuals may be better understood in terms of the probability of achieving one's

maximum storage potential rather than in terms of differences in that potential (Adam et al., 2015b). Likewise, we find that longer retention intervals reduce the probability of achieving one's maximum storage capacity, instead of reducing that storage capacity. Thus, longer retention intervals do not reduce the maximum amount of information that can be stored in working memory, but they provide more opportunity to forget individual items through lapses of attentional control.

CHAPTER 2.

Spatial attention and object-based storage are distinct signatures of working memory

Working memory facilitates the temporary maintenance of small amounts of information so that it can be manipulated or acted upon. Contemporary theories of working memory have coalesced on variations of embedded process models (Cowan, 1999; Oberauer, 2002) in which performance in working memory tasks depends upon memory mechanisms that represent information in two distinct states: an online, active state ("focus of attention"); and an offline, passive state ("silent working memory"). The focus of attention generally refers to information that is currently "in mind", whereas silent working memory is information that was recently within the focus but can still be rapidly accessed. These memory states have been proposed to be implemented at the neural level through persistent neural firing for items within the focus (Curtis & D'Esposito, 2003) and via rapid synaptic plasticity that allows recently attended items to quickly be reinstated (Jonides et al., 2015; Lewis-Peacock et al., 2012; Rose et al., 2016; Stokes, 2015; Wolff et al., 2017).

Here, we seek to characterize the neural mechanisms supporting the focus of attention. Broad neuroscientific support for focus of attention-related activity has been observed in sustained neural firing in monkey electrophysiological studies (Buschman et al., 2011; Funahashi et al., 1993), uni- and multi-variate measurements of BOLD in human fMRI studies (Cowan et al., 2011; Majerus et al., 2016; Todd & Marois, 2004; Xu & Chun, 2006), and sustained electrical and magnetic fluctuations in human EEG and MEG studies (van Dijk et al., 2010; Vogel & Machizawa, 2004a). Within EEG and MEG studies, two candidate measures are consistent with the focus of attention construct. The first is alpha power (8-12hz), which shows sustained modulations during the retention period and has been shown to contain precise spatial

information about the remembered/attended stimulus (Foster et al., 2016; Foster, Bsales, et al., 2017a). Another candidate is the Contralateral Delay Activity (CDA), which is a sustained negativity over the hemisphere contralateral to the positions of to-be-remembered items. CDA amplitude is modulated by the number of items held in working memory, reaches an asymptote once working memory capacity is exhausted, dynamically tracks dropping information, and predicts individual differences in working memory capacity (Unsworth et al., 2014; Vogel et al., 2005; Vogel & Machizawa, 2004a; Williams & Woodman, 2013). A prevailing view of the CDA is that it tracks the number of task-relevant objects that are stored in working memory (Balaban & Luria, 2017a; Luria et al., 2016a).

While the literatures on the CDA and alpha power have largely developed independently, recent proposals claim that they reflect isomorphic measures of the focus of attention. Specifically, (van Dijk et al., 2010) argued that the CDA is an averaging artifact of trial-level alpha modulation, and, therefore, reflects attention to the spatial positions of the memoranda, rather than representations of items in working memory. A similar proposal was made by (Berggren & Eimer, 2016), who found that when two arrays were presented sequentially in different hemifields, CDA amplitude tracked the positions of the most recently seen items (but also see: (Feldmann-Wüstefeld et al., 2018). Such spatial attention accounts make two broad, but untested assertions regarding neural measures of the focus of attention. First, that sustained EEG activity reflecting the focus of attention exclusively represents the current regions of attended space, rather than the online maintenance of the items that occupy those regions of space (Berggren & Eimer, 2016). Second, that such neural measures amount to a monolithic "focus of attention," rather than a collection of distinct, but overlapping mechanisms that together comprise the focus of attention.

Here, we provide evidence that the focus of attention in working memory is not a monolithic construct, but rather, involves at least two neurally separable processes: (1) attention to regions in space (2) representations of objects that occupy the attended regions (i.e., object files). Alpha activity, but not the CDA, tracked attention to relevant spatial positions. Conversely, when participants stored object representations, lateralized alpha activity that tracked the attended positions was accompanied by robust, load sensitive CDA. These results suggest the neural focus of attention can be dissected into at least two complementary, but distinct facets of activity: a map of prioritized space and online representations of objects.

Materials & Methods

Experimental Design

Our broad strategy was to compare delay period activity across two tasks that employed physically identical displays but distinct cognitive requirements. We designed distinct "attention" and "working memory" tasks to disentangle the neural correlates of hypothesized sub-components of the focus of attention. Both tasks are known to recruit sustained spatial attention (i.e. representation of a spatial priority map), but only the working memory task invoked online storage of items (i.e. representation of the objects which occupied the attended locations). For all experiments, participants completed both a working memory task and an attention task, and the sequence of physical stimuli was identical for both tasks; the attention and working memory tasks differed only in the instructions given to participants and in the response mapping to keys. In Experiment 2-1, the working memory task required that participants remember the color of the items in the sample array, whereas the attention condition required participants to direct spatial attention towards the locations of the items in the sample array (item color was irrelevant). Although highly similar, one key difference between the tasks in Experiment 2-1 was that participants were required to remember non-spatial features only in the

working memory task. To test whether the requirement to remember non-spatial features was responsible for our findings in Experiment 2-1, we employed even more similar tasks in Experiment 2-2. The working memory task required that participants *store the spatial positions of items* in the sample array, and the attention task required that participants *covertly attend spatial positions* in anticipation of rare targets during the delay.

Participants

Experimental procedures were approved by the University of Chicago Institutional Review Board. All participants gave informed consent and were compensated for their participation with cash payment ($15 per hour); participants reported normal color vision and normal or corrected-to-normal visual acuity. Participants were recruited from the University of Chicago and surrounding community. For each sub-experiment (e.g., Exp. 2-1a), we set a minimum sample size of 20 subjects (after attrition and artifact rejection). This minimum sample size was chosen to ensure that we would be able to robustly detect set-size dependent delay activity. Prior work employing sample sizes of 10 to 20 subjects per experiment can robustly detect set-size dependent CDA (Vogel et al., 2005; Vogel & Machizawa, 2004a), and differences in CDA amplitude between novel experimental conditions (Balaban & Luria, 2017a). We chose a minimum sample size toward the upper end of this conventional range.

A total of 63 and 54 participants were run in Experiments 2-1 and 2-2, respectively. Due to a technical error, EEG activity was not recorded for 3 participants in Experiment 2-1. In addition, data from some participants was excluded because of excessive EEG artifacts (<120 trials remaining in any of the four experimental conditions) or poor behavioral performance. This left 48 subjects in Experiment 2-1 (28 in Experiment 2-1a, 20 in Experiment 2-1b), and 49 subjects in Experiment 2-2 (20 in Experiment 2-2a, 29 in Experiment 2-2b).

EEG Acquisition

Participants were seated inside an electrically shielded chamber, with their heads resting on a padded chinrest 74 cm from the monitor. We recorded EEG activity from 30 active Ag/AgCl electrodes (Brain Products actiCHamp, Munich, Germany) mounted in an elastic cap positioned according to the International 10-20 system [Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz]. Two additional electrodes were affixed with stickers to the left and right mastoids, and a ground electrode was placed in the elastic cap at position Fpz. Data were referenced online to the right mastoid and re-referenced offline to the algebraic average of the left and right mastoids. Incoming data were filtered [low cut-off = .01 Hz, high cut-off = 80 Hz, slope from low- to high-cutoff = 12 dB/octave] and recorded with a 500 Hz sampling rate. Impedance values were kept below 10 kΩ.

Eye movements and blinks were monitored using electrooculogram (EOG) activity and eye-tracking. We collected EOG data with 5 passive Ag/AgCl electrodes (2 vertical EOG electrodes placed above and below the right eye, 2 horizontal EOG electrodes placed ~1 cm from the outer canthi, and 1 ground electrode placed on the left cheek). We collected eye-tracking data using a desk-mounted EyeLink 1000 Plus eye-tracking camera (SR Research Ltd., Ontario, Canada) sampling at 1,000 Hz. Usable eye-tracking data were acquired for 25 out of 28 participants in Experiment 2-1a, 19 out of 20 participants in Experiment 2-1b, 17 out of 20 participants in Experiment 2-2a, and 29 out of 29 participants in Experiment 2-2b.

Artifact rejection

Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic detection criteria. After automatic detection, trials were manually inspected

to confirm that detection thresholds were working as expected. Subjects were excluded if they had fewer than 120 total trials remaining in any of the 4 conditions. In Experiment 2-1a, we rejected an average 25% of trials across all four conditions. This left us with an average of 282 trials in working memory set size 2 condition, 275 trials in the working memory set size 4 condition, 302 trials in the Attention set size 2 condition, and 302 trials in the Attention set size 4 condition. In Experiment 2-1b, we rejected an average of 32% of trials across all four conditions. This left us with an average of 291 trials in the working memory set size 2 condition, 285 trials in the working memory set size 4 condition, 320 trials in the attention set size 2 condition, and 320 trials in the attention set size 4 condition. In Experiment 2-2a, we rejected an average of 22% of trials across all four conditions. This left us with an average of 302 trials in the working memory set size 2 condition, 301 trials in the working memory set size 4 condition, 322 trials in the attention set size 2 condition, and 323 trials in the attention set size 4 condition. In Experiment 2-2b, we rejected an average of 27% of trials across all four conditions. This left us with an average of 283 trials in the working memory set size 2 condition, 283 trials in the working memory set size 4 condition, 298 trials in the attention set size 2 condition, and 295 trials in the attention set size 4 condition.

Eye movements. We used a sliding window step-function to check for eye movements in the HEOG and the eye-tracking gaze coordinates. For HEOG rejection, we used a split half sliding window approach (window size = 100 ms, step size = 10 ms, threshold = 20 μV). We only used the HEOG rejection if the eye tracking data were bad for that trial epoch. We slid a 100 ms time window in steps of 10 ms from the beginning to the end of the trial. If the change in voltage from the first half to the second half of the window was greater than 20 μV, it was marked as an eye movement and rejected. For eye-tracking rejection, we applied a sliding

window analysis to the x-gaze coordinates and y-gaze coordinates (window size = 100 ms, step size = 10 ms, threshold = 0.5° of visual angle).

Blinks. We used a sliding window step function to check for blinks in the VEOG (window size = 80 ms, step size = 10 ms, threshold = 30 μV). We checked the eye-tracking data for trial segments with missing data-points (no position data is recorded when the eye is closed).

Drift, muscle artifacts, and blocking. We checked for drift (e.g., skin potentials) by comparing the absolute change in voltage from the first quarter of the trial to the last quarter of the trial. If the change in voltage exceeded 100 μV, the trial was rejected for drift. In addition to slow drift, we checked for sudden step-like changes in voltage with a sliding window (window size = 100 ms, step size = 10 ms, threshold = 100 μV). We excluded trials for muscle artifacts if any electrode had peak-to-peak amplitude greater than 200 μV within a 15 ms time window. We excluded trials for blocking if any electrode had at least 30 time-points in any given 200-ms time window that were within 1μV of each other.

Analysis of Horizontal Gaze Position

We rejected all trials that had eye movements greater than 0.5° of visual angle. Nevertheless, participants could still move their eyes within the 0.5° of visual angle threshold (e.g., micro saccades). To compare eye movements in the two tasks, we compared the horizontal gaze position recorded by the eye tracker. We were most concerned with horizontal eye movements, as these could contaminate our lateralized EEG measures. We drift-corrected gaze position data by subtracting the mean gaze position measured 200 ms before the pre-cue to achieve optimal sensitivity to changes in eye position (Cornelissen et al., 2002). We then took the mean change in gaze position (in degrees of visual angle) for left and right trials during same time-window that we used in the CDA analysis, 400 to 1450 ms after stimulus onset. Eye gaze

values from left trials were sign-reversed so that left and right trials could be combined together. As such, positive values indicate eye movements toward the remembered side, and negative values indicate eye movements away from the remembered side. Importantly, not all participants had eye tracking with adequate quality to be included in this analysis. Therefore, only 25 participants from Experiment 2-1a, 19 participants from Experiment 2-1b, 17 participants from Experiment 2-2a, and 27 participants from Experiment 2-2b were included in the analysis.

Analysis of Pupil Dilation

As an additional metric of task difficulty, we compared task-evoked pupil dilation between the working memory and attention tasks. Many studies have demonstrated that task-evoked pupil dilation correlates with cognitive load; the pupil dilates more when there are higher attentional and working memory demands (Beatty, 1982; Steinhauer & Hakerem, 1992). Since we were most interested in assessing the relative difficulty of the two tasks, we collapsed the data across set size within each task. For our analysis, pupil dilation data were baselined from 400 to 0 ms before the onset of the colored squares. Differences in pupil dilation between the working memory and attention tasks (collapsed across set sizes) were calculated by comparing pupil size during the same time-window as is used in the CDA analysis (400 to 1450 ms after stimulus onset). Just as in the analysis of horizontal gaze position, not all participants had eye tracking that was good enough to be included in this analysis. The same participants were included in both the horizontal gaze position and pupil dilation analyses.

Analysis of contralateral delay activity

EEG activity was baselined from 400 ms to 0 ms before the onset of the stimulus array. Trials containing targets for the attention task were excluded. Event-related potentials were calculated by averaging baselined activity at each electrode across all accurate trials within each

condition (Set-Size 2 working memory, Set-Size 4 working memory, Set-Size 2 attention, and Set-Size 4 attention). We calculated amplitude of contralateral and ipsilateral activity for five posterior and parietal pairs of electrodes chosen a priori based on prior literature: O1/O2, PO3/PO4, PO7/PO8, P3/P4, and P7/P8. Statistical analyses were performed on data that was not filtered beyond the .01 – 80 Hz online data acquisition filter; we low-pass filtered data (30 Hz) for illustrative purposes in paper figures.

Analysis of lateralized alpha power

EEG signal processing was performed in MATLAB 2015a (The MathWorks, Natick, MA). We band-pass filtered trial epochs in the alpha band (8-12 Hz) using a bandpass filter from the FieldTrip toolbox (Oostenveld et al., 2011); 'ft_preproc_bandpass.m') and then extracted instantaneous power by applying a Hilbert transform ('hilbert.m') to the filtered data. Trials containing targets for the attention task were excluded. We calculated alpha power for the same five posterior and parietal pairs of electrodes as CDA: O1/O2, PO3/PO4, PO7/PO8, P3/P4, and P7/P8.

Stimuli & Procedures

Stimuli in all experiments were presented on a 24-inch LCD computer screen (BenQ XL2430T; 120 Hz refresh rate) on a Dell Optiplex 9020 computer. Participants were seated with a chinrest 74 cm from the screen. Stimuli were presented on a gray background, and participants fixated a small black dot with a diameter of approximately 0.2 degrees of visual angle.

Experiment 2-1a. We ran two very similar versions of Experiment 2-1 (hereafter referred to as Experiments 2-1a and 2-1b). The stimuli and procedures for Experiments 2-1a and 2-1b were almost identical, with the exception of the differences noted in the "Experiment 2-1b" section. Each trial began with a blank inter-trial interval (750 ms), followed by a diamond cue

(300 ms) indicating the relevant side of the screen (right or left). This diamond cue (0.65° maximum width, 0.65° maximum height) was centered 0.65° above the fixation dot and was half green (RGB = 74, 183, 72) and half pink (RGB = 183, 73, 177). Half of the participants were instructed to attend the green side and the other half were instructed to attend the pink side. After the cue, 2 or 4 colored squares (Exp. 1: 1.1° by 1.1°; Exp. 2: 1° by 1°) briefly appeared in each hemifield (150 ms) with a minimum of 2.10° (1.5 objects) between each item. The squares then disappeared for a 1,300 ms delay period. Squares could appear within a subset of the display subtending 3.1° to the left or right of fixation and 3.5° degrees above and below fixation. Colors for the squares were selected randomly from a set of 9 possible colors (Red = 255 0 0; Green = 0 255 0; Blue = 0 0 255; Yellow = 255 255 0; Magenta = 255 0 255; Cyan = 0 255 255; Orange = 255 128 0; White = 255 255 255; Black = 1 1 1). Colors were chosen without replacement within each hemifield, and colors could be repeated across, but not within, hemifields. On 10% of trials, two small, black lines (0.02° wide, 0.4° long) appeared (66.7 ms), one at the location of a colored square in the cued hemifield, and one at the location of a colored square in the un-cued hemifield. The lines could appear at any point during the delay period from 100 to 1200 ms after the offset of the stimuli. Each line could be tilted 31.3 degrees to the left or 31.3 degrees to the right. At test, a probe display appeared until response, consisting of 1 colored square in each hemifield.

Experiment 2-1b. All stimuli and procedures were the same as Experiment 2-1a with the following exceptions. On 10% of trials, two small, black lines appeared, in the cued hemifield. One line appeared at the location of a colored square. The other line appeared in an unoccupied location in the cued hemifield with a minimum of 2.10° (1.5 objects) from the locations of the memory array items.

Experiment 2-2a. Stimuli were similar to Experiment 2-1b with the following exceptions. Participants were presented with 2 or 4 black circles (0.611° diameter; RGB= 1 1 1) in each hemifield with a minimum of 1.53° degrees (1.5 objects) between each item. These circles and distractor line could appear within a subset of the display subtending 2.44 degrees to the left or right of fixation and 3.06 degrees above and below fixation. On target-present trials, the two small lines that were presented briefly during the retention interval were 0.04° wide and 0.76° long. Both were presented in the attended hemifield. One was at the location of a colored square, and the other was in an unoccupied location that was a minimum of 1.5 objects away from any other memory location.

Experiment 2-2b. Stimuli were similar to Experiment 2-2a with the following exceptions. Participants were presented with 2 or 4 colored circles (0.84° diameter) in each hemifield with a minimum of 2.10° (1.5 objects) between each item. Colors for the circles were randomly selected without replacement within each hemifield from a set of 10 possible colors (Red = 255 0 0; Green = 0 255 0; Blue = 0 0 255; Yellow = 255 255 0; Magenta = 255 0 255; Cyan = 0 255 255; Orange = 255 128 0; Brown = 102 51 0; White = 255 255 255; Black = 1 1 1). On target-present trials, the two small lines that were presented briefly during the retention interval were 0.04° wide and 0.99° long. Both were presented in the attended hemifield. One was at the location of a colored square, and the other was in an unoccupied location that was a minimum of 1.5 objects away from any other memory location.

Participants in all experiments completed a working memory and an attention task (Figure 2-1). Within each experiment, the sequence of physical stimuli was identical for both tasks. Differences in procedures between the experiments are described below. The attention and working memory tasks differed only in the instructions given to participants and in the keys used

to respond. Task order (attention first or working memory first) and relevant cue color (pink or green) were counterbalanced across participants. Participants completed 20 blocks of 80 trials each (1,600 trials total, 400 per condition).

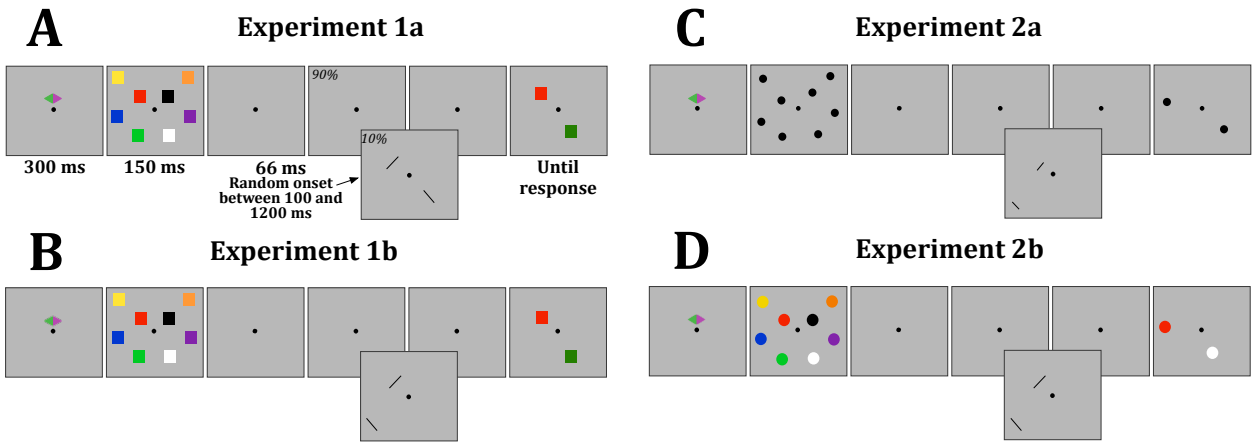

Figure 2-1. Task designs.

Working memory and attention tasks for all experiments. At the start of each trial, a cue appeared on the screen for 300 ms, which cued participants to attend one side of the screen. Then, an array of 2 or 4 colored squares (Exp. 1a & 1b) or circles (Exp. 2a & 2b) briefly appeared (150 ms). On 10% of trials, during the blank retention interval (1300 ms), two small lines appeared for 66 ms between 100 and 1200 ms after memory array offset. In Experiment 2-1a, one line appeared in each hemifield. In all other experiments, both lines appeared in the same hemifield, one in an attended location and one in an unattended location. After the retention interval, a response screen appeared with one square (Exp. 1a & 1b) or circle (Exp. 2a & 2b) in each hemifield. In the working memory task, to respond, participants reported whether the square (Exp. 1a & 1b) or circle (Exp. 2a & 2b) that reappeared in the attended hemifield was the same color (Exp. 1a & 1b) or in the same location (Exp. 2a & 2b). Participants pressed "z" if it was the same color (Exp. 1a & 1b) or location (Exp. 2a & 2b) and "/?" if it was different. In the attention task, if a line was not present during the delay period, participants pressed "spacebar." If a line was present during the delay, participants had to report the orientation of the line that appeared in one of the cued locations. If the line was tilted left, participants pressed "z" and if it was tilted right, participants pressed "/?." The response screen remained visible until a response was made.

Working Memory Task. In Experiments 2-1a and 2-1b, participants were instructed to remember the colors of the presented squares in the cued hemifield and to ignore the lines that might flash during the middle of the delay period. At test, participants were asked to identify whether the color presented at the relevant probed location was the same as the color held in

mind (same trial) or different (change trial). The colors changed on 50% of trials. Participants pressed the "z" key to indicate the response "same" and pressed the "/" key to indicate "different". For Experiment 2-2a and 2-2b, the procedures for the working memory task were very similar to the procedure from Experiment 2-1a and 2-1b, except that participants were asked to identify whether the location of the presented circle in the attended hemifield was in the same or different location as any of the original circles.

Attention Task. Procedures and instructions for the attention task were identical in all experiments. Only the visual stimuli differed, so as to match the visual stimuli presented in the working memory task. Participants were instructed to maintain their attention at the locations of the presented squares in the cued hemifield in order to identify the orientation of a small line that appeared at one of the attended locations on 10% of trials. Participants were instructed to press the "z" key if the line appeared and was tilted left, and the "/" key if the line appeared and was tilted right. On 90% of trials, no line was presented, and participants were instructed to press the "space" key to indicate that there was no target present. The physical stimulus displays were identical to the memory task; thus, one colored square appeared in each hemifield at the end of the attention trials. Participants were told that the appearance of the test display indicated that it was time to respond, and that the location and the color of the squares were irrelevant to the task.

Stimuli and procedures in Experiment 2-1a differed from Experiments 2-1b, 2-2a, and 2-2b only for the target-present trials (10% of trials). Specifically, in Experiments 2-1b, 2-2a, and 2-2b during target-present trials, we presented both a relevant and an irrelevant line within the cued hemifield. One line always appeared at the same location as one of the colored squares; the second line appeared at a foil location where no colored square had been presented (a minimum distance of 1.5 items' width from any of the colored squares' locations). Thus, the participants

were required to maintain their attention at precise locations within the relevant hemifield so that they knew which line to report. We reasoned that the inclusion of an irrelevant item in the cued hemifield in Experiment 2-1b would encourage subjects to orient attention more precisely. However, subsequent analyses revealed no main effect or interactions associated with the changes in procedure between Experiment 2-1a and 2-1b. Therefore, data were collapsed across these two versions of the task.

We would additionally like to note that in both the memory and attention tasks, the circles/squares in the sample array were always at least 1.5 objects apart from each other. In the attention task, the rare target probes appeared at the location of one of the original squares, while the distractor probe appeared in an uncued location that was at least 1.5 objects away from any of the attended locations. Thus, the attention task required subjects to make the same spatial discrimination that subjects had to make in the memory task in order to relate the test probe to the proper item from the memory array. In other words, the positions of the sample items had to be maintained equally precisely in the memory and attention tasks. This is most clear for Experiment 2-2, in which space was the sole relevant attribute for the memory task.

Results

We aggregated data across Experiments 2-1 and 2-2 (n=97) to provide the most power for understanding the distinctions between the working memory and attention tasks. While the aggregate results (Figure 2-2) mirrored those of the individual experiments, the data taken together provide a clear demonstration of the essential empirical patterns. In this aggregate analysis, we focus on CDA, alpha power, and pupil size.

Preliminary analysis of the effect of Experiment

In a preliminary analysis, we examined whether the small variations in task design between Experiments 2-1 and 2-2 had an effect on the observed results. For this purpose, we ran repeated-measures ANOVAs for each analysis (i.e., CDA, alpha power, and pupil size) with the within-subjects factors Task (working memory, attention) and Set Size (2, 4 items) and the between-subjects factor Experiment (2-1, 2-2). For all analyses, there was no main effect of Experiment, $p>=.16$. Therefore, it was justified to collapse data across all experiments.

For the horizontal gaze position and the lateralized alpha analyses, none of the factors significantly interacted with Experiment, $p>=.19$. However, for the pupil dilation analysis, there was a significant interaction of Task and Experiment, $F(1,86)=10.76$, $p=.002$, $\eta p^2 =.11$. This significant interaction is explained by greater pupil dilation in the attention task than in the working memory task in Experiment 2-2, but not in Experiment 2-1.

For the CDA analysis, there was a significant 3-way interaction of Laterality, Set Size, and Experiment, $F(1,95)=6.73$, $p=.01$, $\eta_p^2 =.07$. To further delineate this three-way interaction, we ran follow-up ANOVAs with the factors Laterality (contra, ipsi) and Set Size (2, 4 items) for Experiment 2-1 and Experiment 2-2 separately. These follow-up analyses revealed that there was a significant interaction of Laterality and Set Size for Experiment 2-2, $F(1,48)=16.88$, $p<.001$, $\eta_p^2 =.26$, but not for Experiment 2-1, $F(1,47)=.08$, $p=.78$, $\eta_p^2 =.002$.

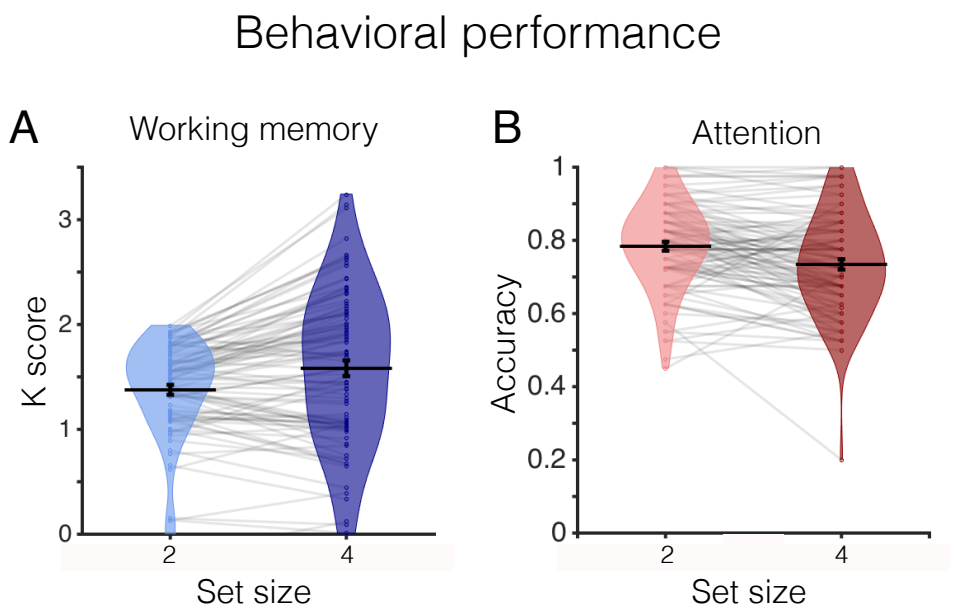

Figure 2-1. Behavioral results.

(A) Average K score in the working memory task for all experiment. The distribution of K scores for all participants is represented by the violin plot. Dots and light gray lines represent one participant's performance. (B) Average accuracy in the attention tasks for all experiment.

Contralateral Delay Activity

Using all data from Experiments 2-1 and 2-2 together, we ran a repeated-measures ANOVA with the factors Task (working memory, attention) and Set Size (2, 4 items). This analysis revealed significant main effects of Laterality, $F(1,96)=74.41$, $p<.001$, $\eta_p^2=.44$, and Set Size, $F(1,96)=33.27$, $p<.001$, $\eta_p^2=.26$. Because these main effects were collapsed across Task, they are not informative for our central question of how storage-related neural signals differ across tasks. Thus, the first important finding was a significant 2-way interaction between Laterality and Task, $F(1,96)=81.27$, $p<.001$, $\eta_p^2=.46$, that reflected a greater laterality effect in the working memory than in the attention task. To confirm this impression, we ran a follow-up 2-way paired-samples t-test that compared contralateral to ipsilateral activity separately for the working memory and attention tasks and each set size (e.g. WM ss2, WM ss4, ATT ss2, ATT

ss4). This analysis revealed that the CDA was significantly more lateralized in the working memory (Set Size 2: M =-.38, SD=.44; Set Size 4: M=-.54, SD=.54) than in the attention (Set Size 2: M=-.09, SD=.34; Set Size 4: M=-.10, SD=.37) task for both set sizes (Set Size 2: $t(98)$=-6.71, p<.001; Set Size 4: $t(98)$=-8.57, p<.001). We note, however, that there was reliable lateralized activity for both tasks, p<=.007.

Another key finding was that the number of items in the sample array had a selective impact on CDA in the working memory task, while CDA in the attention task showed no reliable effect. This impression was verified by a reliable triple interaction between Task, Laterality and Set Size, $F(1,96)=8.75$, p=.004, η_p^2 =.08. As Figure 2-3 shows, CDA was set size dependent in the working memory task, but not in the attention task. To characterize the triple interaction, we ran separate follow-up repeated-measures ANOVAs for each task (working memory and attention) with the factors Laterality (contra, ipsi) and Set Size (2, 4 items). This analysis revealed that there was a significant interaction of Laterality and Set Size for the working memory task, $F(1,96)=14.39$, p<.001, η_p^2 =.13, but not the attention task, $F(1,96)=.07$, p=.79, η_p^2 =.001. Thus, while data from the working memory task showed that CDA amplitude was larger for set size 4 (M=-.55, SD=.54) than set size 2 (M=-.39, SD=.44), data from the attention task showed no evidence of a difference in CDA amplitude between Set Size 2 (M=-.10, SD=.34) and Set Size 4 (M=-.11, SD=.37).

To summarize, the aggregate analysis showed that CDA was substantially stronger in the working memory than in the attention task. Moreover, CDA tracked the increase in mnemonic load from two to four items, while the CDA signal in the attention task – in addition to being over four times smaller than in the working memory task – showed no effect of mnemonic load at all, a defining feature of the CDA. This core result motivates our conclusion that CDA is

directly linked with the online maintenance of object representations in working memory, and not the deployment of attention to the positions of the sample items.

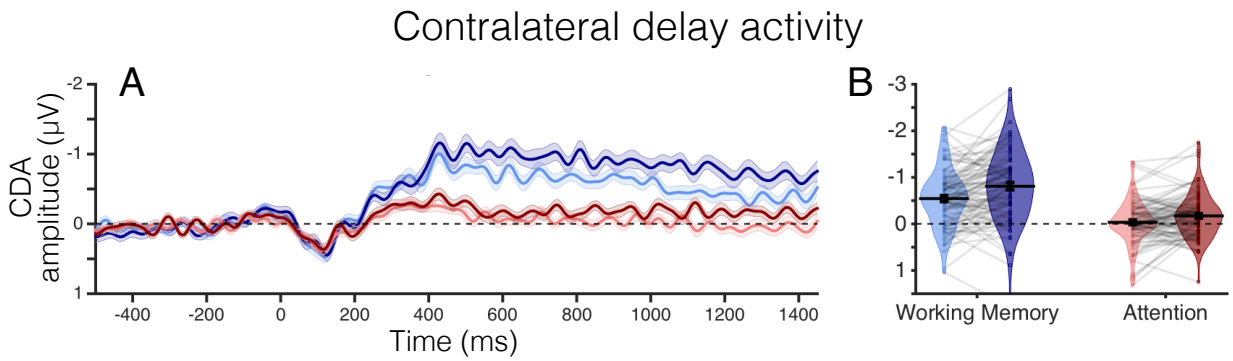

Figure 2-2. Contralateral delay activity results.

Contralateral delay activity. (A) CDA amplitude (μV) over time for all experiments combined. Timepoint zero marks the onset of the memory array, and timepoint 1450 marks the onset of the response array. (B) Average CDA amplitude (μV) for all experiments during the time window of interest, 400 to 1450 ms. The distribution of CDA amplitudes for all participants is represented by the violin plot. Dots and light gray lines represent one participant's CDA amplitude.

Lateralized Alpha Power

As Figure 2-4 shows, we observed the typical suppression of alpha power contralateral to the relevant hemifield in both tasks, though it was larger in the working memory task. We confirmed these impressions with a repeated-measures ANOVA on the average alpha power with the factors Laterality (contra, ipsi), Task (attention, working memory) and Set Size (2, 4 items). This analysis revealed a significant main effect of Laterality, $F(1,96)=45.57$, $p<.001$, $\eta_p^2=.32$, and a significant interaction between Laterality and Set Size, $F(1,96)=9.75$, $p=.002$, $\eta_p^2=.09$. Paired t-tests confirmed that this interaction reflects a stronger lateralization of alpha power in the set size 2 condition (M=-12.24, SD=17.17) than in the set size 4 condition (M=-10.04, SD=16.06) ($t(96)=-3.123$, $p=.002$). Thus, the strength of lateralized alpha activity varied with the number of stored or attended positions in both the working memory and attention tasks. Critically, however, the effect of set size on lateralized alpha power was in the opposite direction

from the effect we observed with CDA. CDA was stronger for set size 4 than for set size 2 whereas alpha lateralization was stronger for set size 2 than for set size 4. These findings support the hypothesis that CDA and alpha activity reflect distinct aspects of online storage in visual working memory.

Our analysis also revealed a significant interaction between Laterality and Task, $F(1,96)=27.22$, $p<.001$, $\eta_p^2=.22$, that reflected the greater lateralization of alpha power in the working memory than in the attention task. This impression was confirmed with a two-way paired samples t-test that revealed a significant difference in alpha power lateralization between the working memory (M=-7.79, SD=11.61) and attention (M=-3.35, SD=5.69) tasks, $t(96)=-5.22$, $p<.001$. Critically, both tasks showed clear evidence of lateralized alpha power in both set sizes ($p<.001$ for all conditions), confirming that covert attention was deployed to the position of sample items in a sustained fashion in both tasks. The greater lateralization of alpha power in the working memory than attention task is a robust empirical pattern that is present in both experiments and in the aggregate analysis. Though we did not expect this pattern *a priori*, this reliable difference in alpha lateralization between the two tasks may reflect a direct influence of online object representations on the deployment of spatial attention.

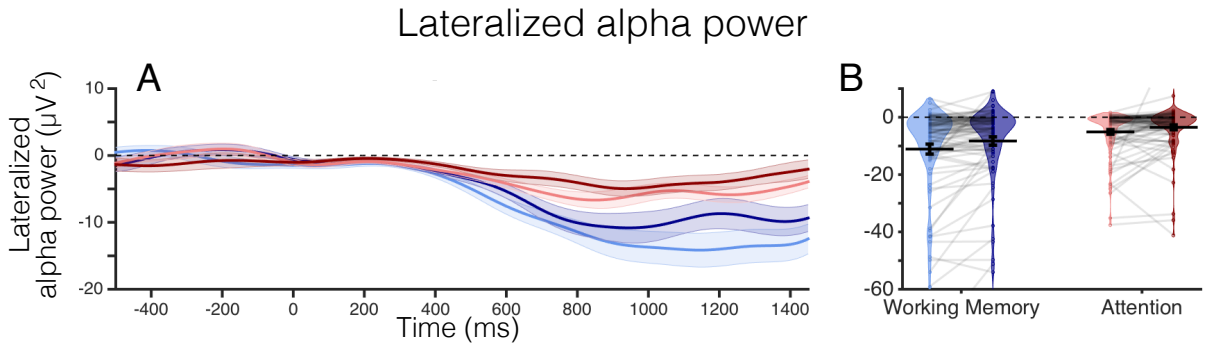

Figure 2-3. Lateralized alpha power results.

(A) Lateralized alpha power (μV^2) over time collapsed across all experiments. Timepoint zero marks the onset of the memory array, and timepoint 1450 marks the onset of the response array.

(Figure 2-4 continued) (B) Average lateralized alpha power (μV^2) for all experiments during the time window of interest, 400 to 1450 ms. The distribution of lateralized alpha power for all participants is represented by the violin plot. Dots and light gray lines represent one participant's alpha power.

Pupil Dilation

We argue that the working memory task encouraged online storage of object representations while the attention task did not. Thus, the restriction of load dependent CDA to the working memory task could reflect a direct link between the CDA and item storage in working memory. A clear alternative hypothesis, however, is that the working memory task may differ from the attention task in terms of the intensity or effort applied to the task rather than the specific cognitive operations that were invoked. While accuracy was similar (and off of ceiling) in the two tasks, this does not provide strong evidence for equivalent effort. Fortunately, pupil dilation measurements have been shown to provide a sensitive index of cognitive effort and arousal when bottom-up stimulus factors are controlled. Thus, we ran a two-way paired-samples t-test to examine whether pupil size differed during the time window in which CDA was measured. This analysis revealed a greater level of pupil dilation in the attention task (M= .63, SD=1.89) compared to the working memory task (M= .29, SD=2.20), $t(87)=-3.13$, $p=.002$, suggesting that the attention task recruited greater levels of cognitive effort. Thus, our finding that CDA was far larger in the working memory task cannot be explained by increased effort in the working memory task. Indeed, pupil analysis of the aggregated data suggests that the working memory task was the easier of the two. These findings argue for a difference in the nature of the cognitive operations evoked by the working memory and attention tasks, rather than in the degree to which similar operations were carried out.

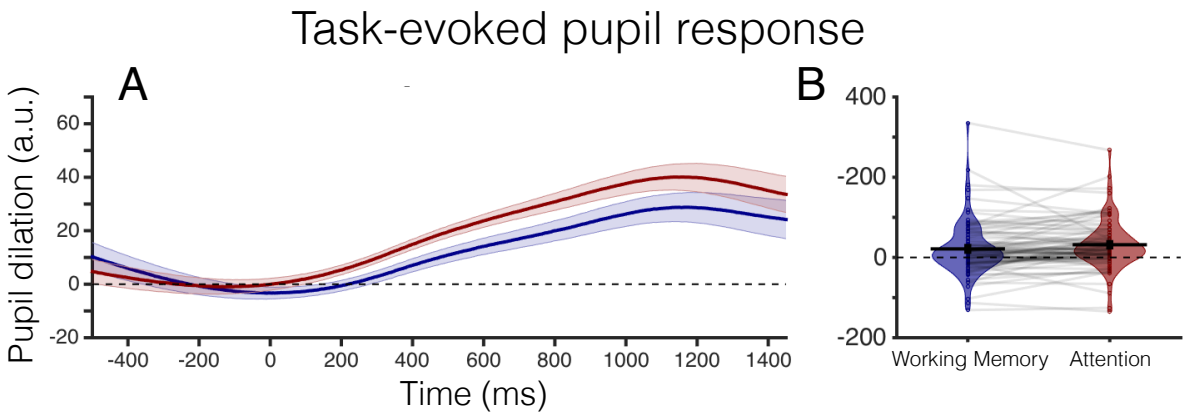

Figure 2-4. Task-evoked pupillary response.

(A) Pupil dilation (arbitrary units) over time for all experiments combined. Timepoint zero marks the onset of the memory array, and timepoint 1450 marks the onset of the response array. (B) Average pupil dilation (arbitrary units) for all experiments during the time window of interest, 400 to 1450 ms. The distribution of pupil dilation for all participants is represented by the violin plot. Dots and light gray lines represent each participant's pupil dilation.

Summary

With 97 subjects, our aggregate analysis provided strong statistical power for documenting how neural activity differed between the working memory and attention tasks. CDA was more than four times larger in the working memory task than in the attention task. Moreover, CDA in the working memory task clearly tracked changes in mnemonic load whereas CDA in the attention task showed no evidence of load sensitivity. Thus, given that these tasks employed identical stimulus displays, we conclude that CDA may be directly tied to the unique *object representation* requirements in the working memory task and not covert attentional orienting to the sample array positions.

The working memory and attention tasks differ in terms of object storage, but past work suggests that both working memory and attention tasks may call upon a common spatial attention process that elicits orderly changes in the scalp topography of alpha power. In addition to past studies showing the broad involvement of alpha activity across a wide range of attention and

memory paradigms (Canolty & Knight, 2010; Fries, 2005; Klimesch, 2012), more recent work has also established that the topography of alpha activity on the scalp can be used to precisely track the locus of covert attention (Foster, Sutterer, et al., 2017; Rihs et al., 2007) and locations stored in working memory (Foster et al., 2016; Foster, Bsales, et al., 2017a). In line with this work, there was clear evidence from both the working memory and attention tasks that alpha power in posterior electrodes was reduced contralateral to the sample array. Importantly, the aggregate analysis also had enough power to reveal a reliable effect of set size on the strength of alpha lateralization, such that greater lateralization was observed in the set size 2 condition compared to the set size 4 condition. This effect should be interpreted with caution, however, as some previous research has not found an effect of set size on lateralization (Fukuda et al., 2015), while others have found greater lateralization for larger set sizes (Sauseng et al., 2009). Nevertheless, the fact that CDA showed the opposite pattern, with higher CDA for the larger set size, highlights the possibility that these two neural signals (measured from within the same set of electrodes) index distinct aspects of maintenance within the focus of attention.

There were differences in probe probability between the attention and working memory tasks. The working memory task required participants to make an attentionally demanding response after every trial. However, in the attention task, participants only had to make an attentionally demanding orientation discriminate on 10% of trials. On the remaining 90% of trials, participants had to press space, indicating that no lines were present. To our knowledge, the influence of probe probability on CDA amplitude has not been investigated. Therefore, the difference in CDA amplitude in the two tasks could be affected by probe probability. We think this alternate explanation is unlikely, as a large body of research has shown that the CDA tracks information stored in working memory across a wide range of response modalities, including

two-alternative change detection (Vogel & Machizawa, 2004a), whole report of discrete colors (Adam et al., 2018), and tracking of dynamic displays (Balaban & Luria, 2017a; Drew & Vogel, 2008).

We additionally examined whether the observed differences in neural activity were a consequence of differential effort or arousal in the two tasks. Because the stimulus displays were identical, we were able to use task-evoked pupil dilation to obtain a sensitive metric of cognitive effort and arousal. The aggregate analysis revealed that the attention task elicited reliably larger pupil size than the working memory task, suggesting that the attention task elicited greater effort. In line with this conclusion, we also note that while behavioral data from the attention task showed that monitoring four locations was more difficult than monitoring two locations, CDA in the attention task was unaffected by set size. Therefore, our findings argue strongly against the hypothesis that stronger delay period signals in the working memory task were a consequence of greater cognitive effort.

Discussion

The focus of attention refers to the small set of mental representations that can be held in an *online* or readily accessible state. Motivated by its central role in intelligent behaviors, there has been a longstanding effort to elucidate the neural signals that track the contents of this internally attended information. This body of work has tended to treat the focus as a monolithic entity, and has claimed that internal attentional processes influence selection and maintenance of cognitive representations in the absence of sensory input (Chun, 2011; Chun et al., 2011). This idea has been supported by neural evidence that has found that sustained working memory representations in the brain occur in the same regions as perceptual representations, which are inherently modulated by attentional mechanisms (Postle, 2006). However, in this study, we

extend the growing evidence that the focus of attention may be implemented via multiple component processes playing distinct functional roles: one that represents currently prioritized space (alpha); and another that reflects item storage within the focus of attention (CDA). This proposal converges with other findings that suggest a dissociation between spatial attention and working memory storage (Tas et al., 2016).

Contralateral delay activity and lateralized alpha power: Distinct components of the focus of attention

van Dijk et al., 2010 proposed that asymmetric modulations of alpha power at the trial-level can generate a CDA-like negative slow wave in an event-related average. However, there is growing evidence that these two measures can be clearly dissociated. For example, (Fukuda et al., 2016) used a lateralized change detection task where they cued participants to one side of the screen, but had a longer than normal (1,000 ms) SOA between the cue and the memory array. During this blank cue period, participants knew which hemifield would contain memory items, but no items had yet appeared. During this time, there was robust alpha power lateralization but no CDA. However, after the memory array appeared, the CDA and alpha power lateralization appeared in concert during the memory maintenance period (1,000 ms). These results suggest lateralized alpha power, and thus attention, can be shifted to empty space, but that the CDA necessitates object storage (see also Fukuda et al., 2015). A similar dissociation between attentional deployment to objects and to spatial location has been found with the anticipatory N2PC, a component which is related the CDA (Woodman et al., 2009).

Contralateral delay activity as an index of item-based storage in working memory

What was the critical difference between the working memory and attention tasks? Despite the fact that they employed identical stimulus displays, the amplitude of the CDA was more than four times larger in the working memory than in the attention task, and only the working memory task elicited load dependent CDA. Both tasks elicited covert orienting to the positions of the items in the sample array, as shown by sustained lateralized alpha power modulations. Moreover, despite distinct monikers, *both* tasks required the sustained maintenance of spatial information across a blank delay. This storage requirement is obvious for the working memory change detection task. But even for the attention task, subjects must have maintained the cued positions so that they could distinguish targets from lures. Indeed, in all experiments, change detection in the working memory task required precisely the same spatial discriminations as did target identification in the attention task. Thus, we propose that the critical difference between the working memory and attention tasks was that the working memory task encouraged the continued *representation of the items* in the sample array, while in the attention task participants directed spatial attention to those positions without maintaining the items themselves.

Contralateral delay activity as a neural index of object file maintenance

Our interpretation of the CDA as an index of continued representations of object files critically hinges on a distinction between the maintenance of items in working memory and the maintenance of spatial information without an accompanying item representation. While some may view this as provocative, recent work has shown dissociable patterns of activity in parietal lobe between working memory and spatial attention demands (Sheremata et al., 2018). Additionally, we note that there is a longstanding precedent for a distinction between the

representation of an object and the representation of the features or identifying labels associated with that object. (Kahneman et al., 1992) elucidated this idea with the *object file* construct which proposes two separable stages of processing. The first involves the parsing of the scene into a set of individuated items that are indexed based on their spatial and temporal coordinates. Subsequently, the specific feature values (e.g., color and orientation) are processed and incorporated into the associated object file. Thus, object files anchor the episodic representation in a specific time and place and are distinct from the specific feature values that are bound together by virtue of an object file. In the present context, an intriguing possibility is that CDA indexes the maintenance of object files in working memory. This proposal is consistent with recent work showing that the CDA is sensitive to objecthood cues (Balaban & Luria, 2016) and tracks the number of encoded objects, not the number of features within objects (Luria & Vogel, 2011). Thus, even though the attention task required the sustained maintenance of location information, CDA was minimal or absent (and insensitive to mnemonic load) because the task did not encourage the maintenance of the object files that were created during the encoding of the sample array. Finally, while object files have been argued to mediate the binding of multiple features within an object, we note that this does not preclude the operation of object files for single-feature objects (Kahneman et al., 1992), such as those required by the spatial working memory task of Experiment 2-2. Thus, we propose that the CDA may provide a neural index of object file maintenance.

Open question on the impact of "object files" on the allocation of spatial attention

In this series of experiments, lateralized alpha power was a useful tool to illustrate that participants sustained their attention to the cued side even when the CDA was completely absent

(Exp 1). However, we also observed a main effect of our task manipulation on lateralized alpha power. When task demands required participants to encode object representations, alpha power was significantly more lateralized than when they only had to sustain their attention to empty space. Though we did not predict this pattern *a priori*, it was reliable in both experiments. This suggests that, like the CDA, lateralized alpha power respects the dissociation between forming object representations and maintaining a spatial priority map. One possible interpretation of this effect is that object representations serve as "anchors" for the allocation of spatial attention, thus amplifying the effects of attention and leading to increased alpha power lateralization. Indeed, such an anchoring effect may provide a productive perspective on prior demonstrations of object-based attention (Egly et al., 1994). While future work is needed to investigate the complex interrelationship between lateralized alpha power and online object representations, the present work clearly suggests that lateralized alpha power does not directly generate, and is dissociable from, the CDA.

Conclusions

A growing body of evidence has shown that CDA and alpha power are tightly linked with the maintenance of information in the focus of attention. Here, we present new evidence that these two neural signals represent distinct facets of this online system. A topographic distribution of alpha power indexes the current locus of spatial attention, a process that is integral to both visual selection and the voluntary storage of items in working memory. By contrast, CDA tracks the active maintenance of object files, the item-based representations that allow observers to integrate the ensemble of features and labels that are associated with visual objects. The dissociable activity of the CDA and alpha power suggests that the focus of attention is composed

of at least two distinct but complementary neural processes, a conclusion with strong implications for both cognitive and neural models of this online storage system.

PART II.

Interrupting working memory

CHAPTER 3.

Spatial attention and object-based storage distinctly respond to interruptions

Working memory is a large-scale neural system that maintains readily accessible task-relevant information via active neural firing. A key challenge for this system is to protect these active representations from task-irrelevant interruptions. Extensive prior work has characterized how the presence of irrelevant information during the encoding of targets (distractors) impacts working memory representations (Clapp et al., 2010; Feldmann-Wüstefeld & Vogel, 2018; Gaspar & McDonald, 2014; Postle et al., 2005). This work has revealed that the presence of distractors during this initial encoding period (0-500ms) greatly reduces working memory performance, in part because these items compete with targets for limited representational space in working memory (De Fockert et al., 2001; Olivers, 2008; Vogel et al., 2005). After this initial encoding period, presence of irrelevant information (interrupters) has a reduced, but still measurable impact on performance (Vogel et al., 2006). These interrupters have less of an impact because working memory representations have reached a more stable state, which is consistent with the time-course of neural measures of working memory representations (Ikkai et al., 2010; Vogel et al., 2006). This reduced impact is also likely due to the formation of concurrent visual long-term memory representations that represent the targets in a passive, yet still accessible format (Chun & Turk-Browne, 2007; Woodman & Chun, 2006). Together, these concurrent active and passive representations of targets reduce the behavioral impact of interruption during working memory maintenance. Yet, despite robust behavioral performance, current models of working memory still predict that onsets of task-irrelevant interruption should produce a momentary perturbation of the maintained target representations during which

attention is withdrawn from the targets and at least temporarily applied to the positions of the interrupters (Bisley et al., 2004; Bisley & Goldberg, 2003). However, the consequences of such a brief withdrawal of attention on the neural signatures of working memory are not well understood. Here, we seek to measure the impact that task-irrelevant interruption has on the ongoing active neural representations of targets held in working memory.

To track the impact of task-irrelevant interruption on neural representations of working memory, we measured two well-characterized, temporally sensitive EEG markers that reflect active, prioritized working memory representations: the contralateral delay activity (CDA) and lateralized alpha power (8-12hz). The CDA is a sustained negative-going wave in human EEG that tracks current working memory load. It is sensitive to trial-by-trial fluctuations in working memory performance and distinguishes stable individual differences in working memory (Vogel and Machizawa, 2004; Luria *et al*., 2016). This component is thought to reflect an index of the current items that are actively represented in working memory (Feldmann-Wüstefeld, Vogel and Awh, 2018; Hakim *et al*., 2018). Lateralized alpha power is similarly sensitive to task-relevant information. This signal is measured as a decrease in alpha power over posterior electrodes that are contralateral to the position of the attended items. However, despite its similarity to the CDA, it has been shown to be a distinct component of actively maintained information (Fukuda et al., 2015) that appears to primarily track the current position of spatial attention (Worden *et al*., 2000; Thut *et al*., 2006; Foster *et al*., 2016, 2017). Topographic distributions of alpha power across the entire scalp have been shown to contain precise spatial information about remembered/attended stimuli (Foster *et al*., 2017; van Moorselaar *et al*., 2017), whereas lateralized alpha power have been used as an effective tool for establishing which visual hemifield is currently attended. Together, the CDA and lateralized alpha power respectively

provide an item-based and space-based index of task-relevant information that is actively represented in working memory. Furthermore, because both signals are lateralized, we were able to isolate processing of the memory array by presenting the memory items laterally and the interrupters along the vertical midline of the display. As items on the vertical midline do not affect lateralized signals, the ongoing lateral measures only reflect processing of the memory representations. This allowed us to measure how these working memory representations respond to task-irrelevant interruption.

Overview of Experiments

In Experiment 3-1, we sought to determine how task-irrelevant interrupters impact ongoing working memory representations. We did this by presenting midline interrupters during the retention interval of a working memory task. In Experiment 3-2, we sought to determine whether the neural responses to task-irrelevant interrupters could be modulated by task expectancy. During all of these tasks, we recorded EEG activity from human participants.

Experiment 3-1

Materials and Methods.

Participants. Twenty-two volunteers, naïve to the objective of the experiment participated for payment (~15 USD per hour). All data were collected in a single session. The data of 2 participants were excluded from the analysis because of too many artifacts, poor behavioral performance (see below for criteria) or technical problems. The remaining 20 participants (12 male) were between the ages of 19-30 (M = 22.7, SD = 3.4). Participants in all experiments reported normal or corrected-to-normal visual acuity as well as normal color vision.

All experiments were conducted with the written understanding and consent of each participant. The University of Chicago Institutional Review Board approved experimental procedures.

Stimuli. All stimuli were presented on a gray background (~33.3 cd/m^2). Cue displays showed a small central fixation dot (0.2° × 0.2°). A horizontal diamond comprised of a green (RGB = 74, 183, 72; 52.8 cd/m^2) and a pink (RGB = 183, 73, 177; 31.7 cd/m^2) triangle appeared on the vertical midline 0.65° above the fixation dot. In 50% of the trials, the pink triangle pointed to the left side and the green triangle pointed to the right side, in the remaining 50% of the trials this was inverse. Half the participants were instructed to attend the hemifield that the pink triangle pointed to, and the other half was instructed to attend the hemifield the green triangle pointed to. Memory displays showed a series of colored squares (1.1° by 1.1°, mean luminance 43.1 cd/m^2). Colors for the squares were selected randomly from a set of 11 possible colors (Red = 255 0 0; Green = 0 255 0; Blue = 0 0 255; Yellow = 255 255 0; Magenta = 255 0 255; Cyan = 0 255 255; Purple = 102 0 102; Brown = 102 51 0; Orange = 255 128 0; White = 255 255 255; Black = 0 0 0). Squares could appear within an area of the display subtending 3.5° to the left or right of fixation and 3.1° above and below fixation. There was the same number of squares in each hemisphere. Within each hemisphere, squares were as equally distributed between the upper and lower hemifields as possible. The interruption display showed four colored squares of the same size as the memory display, drawn from the remaining colors. These interrupting items were shown on the vertical midline with a randomly jittered horizontal offset of maximally 0.55° (half of an object). Retention interval displays were blank with a small central fixation dot (0.2° × 0.2°). Probe displays showed one colored square in each hemisphere in the same location as one of the squares in the original array. In 50% of the trials, the color was identical (no change trials) to the memory display. In the remaining 50% of the trials, it was one of the colors not used

in the memory or interruption display (change trials). The same stimuli were used in all experiments.

Apparatus. Participants were seated with a chinrest in a comfortable chair in a dimly lit, electrically shielded and sound attenuated chamber. Participants responded with button presses on a standard keyboard that was placed in front of them. Stimuli were presented on an LCD computer screen (BenQ XL2430T; 120 Hz refresh rate; 61 cm screen size in diameter; 1920 × 1080 pixels) placed at 74 cm distance from participants. An IBM-compatible computer (Dell Optiplex 9020) controlled stimulus presentation and response collection.

Procedure. Each trial began with a cue display (500 ms) indicating the relevant side of the screen (left or right). A memory display consisting of 6 colored squares in each hemisphere followed the cue display for 150 ms. Participants were instructed to memorize as many colored squares in the memory display from the cued side and to ignore the other side entirely, as that side would never be probed. Participants had to remember the items for a retention interval of 1,650 ms during which a central fixation dot was shown. In 25% of the trials, an interruption display appeared 500 ms after memory display offset for 150 ms. The total length of the retention interval was 1,650 ms, regardless of whether an interruption appeared. Participants were instructed to always ignore interruption displays. After the retention interval, a probe display appeared until response. Participants had to indicate whether the color at the probed location changed color ("?/" key) or did not change color ("z" key). After participants responded, the trial concluded, and the next trial started after a blank inter-trial interval of 750 ms. Participants completed a total of 1200 trials (15 blocks of 80 trials), i.e., 300 trials with interruption and 900 trials without interruption. Information about average performance and a

minimum break of 30 seconds was provided after each block. See Figure 3-1 for a visual depiction of the task.

We presented the interrupters in locations that did not overlap with the locations of the memory items to avoid visual masking. Importantly, the relative position of interrupters and targets matter in lateralized change detection tasks. When interrupters are presented laterally with targets on the vertical midline, the neural signature of sustained interrupters suppression can be isolated (CDAp). Conversely, when interrupters are presented on the vertical midline and targets are presented laterally, the neural signature of target processing can be isolated (Feldmann-Wüstefeld & Vogel, 2018). Accordingly, since we were interested in how neural representations of targets are affected by interruption, we placed the interrupters along the vertical midline. Thus, reductions in CDA amplitude can be interpreted as dropping memory items, and reductions in lateralized alpha power can be interpreted as a shift of attention away from the laterally presented memory arrays.

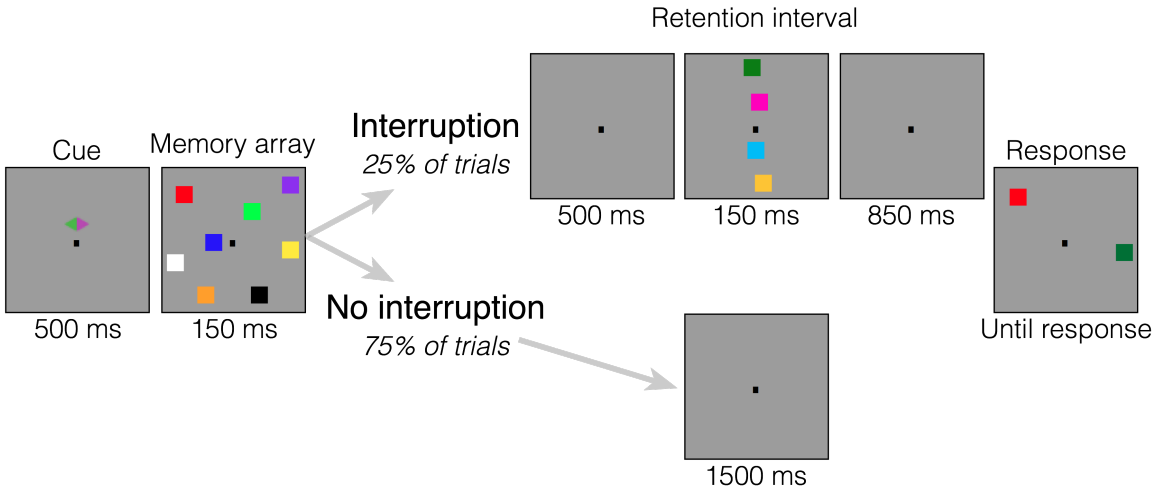

Figure 3-1. Task design.

At the start of each trial, a cue appeared on the screen for 500 ms, which cued participants to attend one side of the screen. Then, an array of 4 colored squares briefly appeared (150 ms). On 75% of trials (no interruption condition), the retention interval (1500 ms) remained blank the

(Figure 3-1 continued) entire time. On the other 25% of trials (interruption condition), the retention interval was blank for 500 ms, but then a series of four colored squares appeared on the midline for 150 ms. Participants were told to always ignore these squares that appeared on the midline of the screen during the retention interval. Following the brief interruption, the screen then went blank again for 850 ms. At the end of each trial, a response screen appeared with one square in each hemifield. Participants were told to report whether the square on the attended side was the same color as the original square in that location.

Behavioral Data Analysis.

We separately analyzed performance for the trials with and without interruption. Performance was converted to a capacity score, K, calculated as N x (H-FA), where N is the set-size, H is the hit rate, and FA is the false alarm rate (Cowan et al., 2011). To compare performance between the two conditions, we used a two-tailed, repeated measures t-test.

Artifact Rejection.

We recorded EEG activity from 30 active Ag/AgCl electrodes (Brain Products actiCHamp, Munich, Germany) mounted in an elastic cap positioned according to the International 10-20 system [Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz]. FPz served as the ground electrode and all electrodes were referenced online to TP10 and re-referenced off-line to the average of all electrodes. Incoming data were filtered [low cut-off = .01 Hz, high cut-off = 80 Hz, slope from low- to high-cutoff = 12 dB/octave] and recorded with a 500 Hz sampling rate. Impedances were kept below 10kΩ. To identify trials that were contaminated with eye movements and blinks, we used electrooculogram (EOG) activity and eye tracking. We collected EOG data with 5 passive Ag/AgCl electrodes (2 vertical EOG electrodes placed above and below the right eye, 2 horizontal EOG electrodes placed ~1 cm from the outer canthi, and 1 ground electrode placed on the left cheek). We collected eye-tracking data using a desk-mounted EyeLink 1000 Plus eye-tracking camera (SR Research Ltd., Ontario, Canada) sampling at 1,000

Hz. Usable eye-tracking data were acquired for 20 out of 22 participants in Experiment 3-1 and 29 out of 30 participants in Experiment 3-2.

EEG was segmented off-line with segments time-locked to memory display onset. Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic detection criteria to each segment. After automatic detection (see below), trials were manually inspected to confirm that detection thresholds were working as expected. Participants were excluded if they had less than 100 correct trials remaining in any of the conditions. For the participants used in analyses, we rejected on average 21% of trials in Experiment 3-1, and 39% of trials in Experiment 3-2.

Eye movements. We used a sliding window step-function to check for eye movements in the HEOG and the eye-tracking gaze coordinates. For HEOG rejection, we used a split half sliding window approach. We slid a 100 ms time window in steps of 10 ms from the beginning to the end of the trial. If the change in voltage from the first half to the second half of the window was greater than 20 μV, it was marked as an eye movement and rejected. For eye-tracking rejection, we applied a sliding window analysis to the x-gaze coordinates and y-gaze coordinates (window size = 100 ms, step size = 10 ms, threshold = 0.5° of visual angle).

Blinks. We used a sliding window step function to check for blinks in the VEOG (window size = 80 ms, step size = 10 ms, threshold = 30 μV). We checked the eye-tracking data for trial segments with missing data-points (no position data is recorded when the eye is closed).

Drift, muscle artifacts, and blocking. We checked for drift (e.g., skin potentials) by comparing the absolute change in voltage from the first quarter of the trial to the last quarter of the trial. If the change in voltage exceeded 100 μV, the trial was rejected for drift. In addition to slow drift, we checked for sudden step-like changes in voltage with a sliding window (window

size = 100 ms, step size = 10 ms, threshold = 100 μV). We excluded trials for muscle artifacts if any electrode had peak-to-peak amplitude greater than 200 μV within a 15 ms time window. We excluded trials for blocking if any electrode had at least 30 time-points in any given 200-ms time window that were within 1V of each other.

Contralateral Delay Activity Analysis

Segmented EEG data was baselined from 200 ms to 0 ms before the onset of the memory displays. Artifact-free EEG segments from correct trials were averaged separately for each condition (no interruption, interruption) and separately for electrodes ipsi- and electrodes contralateral to the attended side. Then the difference between contra- and ipsi-lateral activity for the electrode pair PO7/PO8 was calculated (i.e., the CDA), resulting in two average waveforms for each participant. The average CDA amplitude was calculated for three time windows: before interruption onset (450-650 ms), after interruption onset (800-1,000 ms), before probe onset (1,300-1,500 ms). We then compared the CDA for each time window with a repeated measures two-tailed t-test. To measure the robustness of the CDA for each condition (reliable difference between contra- and ipsilateral activity), we also ran a one-sample t-test (against zero) for each time window.

Lateralized Alpha Power Analysis

The same EEG segments as the CDA analysis were used in this analysis, however, the segments were not baselined. The raw EEG signal was band-pass filtered in the alpha band (8-12 Hz) using a two-way least-squares finite-impulse-response filter ("eegfilt.m" from EEGLAB Toolbox). Instantaneous power was then extracted by applying a Hilbert transform ('hilbert.m') to the filtered data. The resulting data were averaged separately for each condition (no interruption, interruption) and each laterality (contra- versus ipsi-lateral to cued hemifield) for

the electrode pair PO7/PO8, resulting in four average waveforms for each participant. The average alpha power was calculated for the same three time windows as the CDA analysis. We then compared lateralized alpha power suppression for each time window with a repeated-measures two-tailed t-test. To measure the robustness of lateralized alpha power suppression for each condition (reliable difference between contra- and ipsilateral activity), we also ran a one-sample t-test (against zero) for each time window.

Results

Behavior

Performance (Figure 3-2), as measured by K, was significantly worse on trials that were interrupted (M=1.6), than on trials that were not interrupted (M=1.2), significant two-way repeated measures t-test, $t(19)=4.428$, $p<0.001$, $M=0.408$, 95% CI (0.215, 0.601).

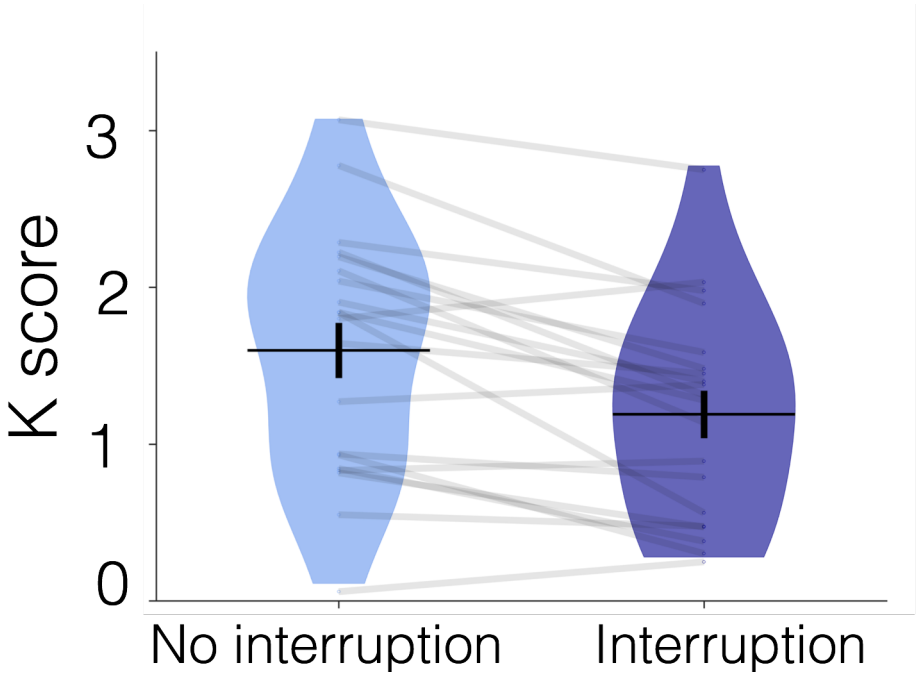

Figure 3-2. Behavioral performance for Experiment 3-1.

(Figure 3-2 continued) Average performance is represented with the horizontal black line. The standard error is represented as the vertical black line. The distribution of K scores for all participants is represented by the violin plot. Dots and light gray lines represent one participant's performance.

Contralateral Delay Activity

Pre-Interruption (450-650 ms). The CDA (Figure 3-3) was robust before interruption onset (450-650 ms) on trials with, $t(19)=-3.187$, $p=0.005$, $M=-0.707$, 95% CI (-1.171 -0.243), and without, $t(19)=-4.053$, $p=0.001$, $M=-0.837$, 95% CI (-1.270 -0.405) interruption. There was not a significant difference in CDA amplitude between trials with and without interruption during this time window, $t(19)=-1.394$, $p=0.179$, $M=-0.131$, 95% CI (-0.327, 0.066).

Post-Interruption I (800-1,000 ms). Immediately following the offset of the interruption (800-1,000 ms), the CDA remained robust for both conditions, No Interruption: $t(19)=-4.016$, $p=0.001$, $M= -0.674$, 95% CI (-1.025 -0.323); Interruption: $t(19)=-2.928$, $p=0.009$, $M=-0.583$, 95% CI (-1.000 -0.166). Again, there was not a significant difference in CDA amplitude between trials with and without interruption, $t(19)= -0.525$, $p= 0.606$, $M=-0.091$, 95% CI (-0.452 0.271).

Post-interruption II (1,300-1,500 ms). By the end of the trial (1,300-1,500 ms), however, there was a significant difference in CDA amplitude between trials with and without interruption, $t(19)= -5.145$, $p= <0.001$, $M=-0.731$, 95% CI (-1.028, -0.434). On trials without interruption, the CDA remained robust, $t(19)=-2.535$, $p=0.020$, $M= -0.496$, 95% CI (-0.906 -0.086) . However, on trials with interruption, the CDA was no longer significantly different from zero, 1.445, $p=0.165$, $M=0.234$, 95% CI (-0.105 0.574).

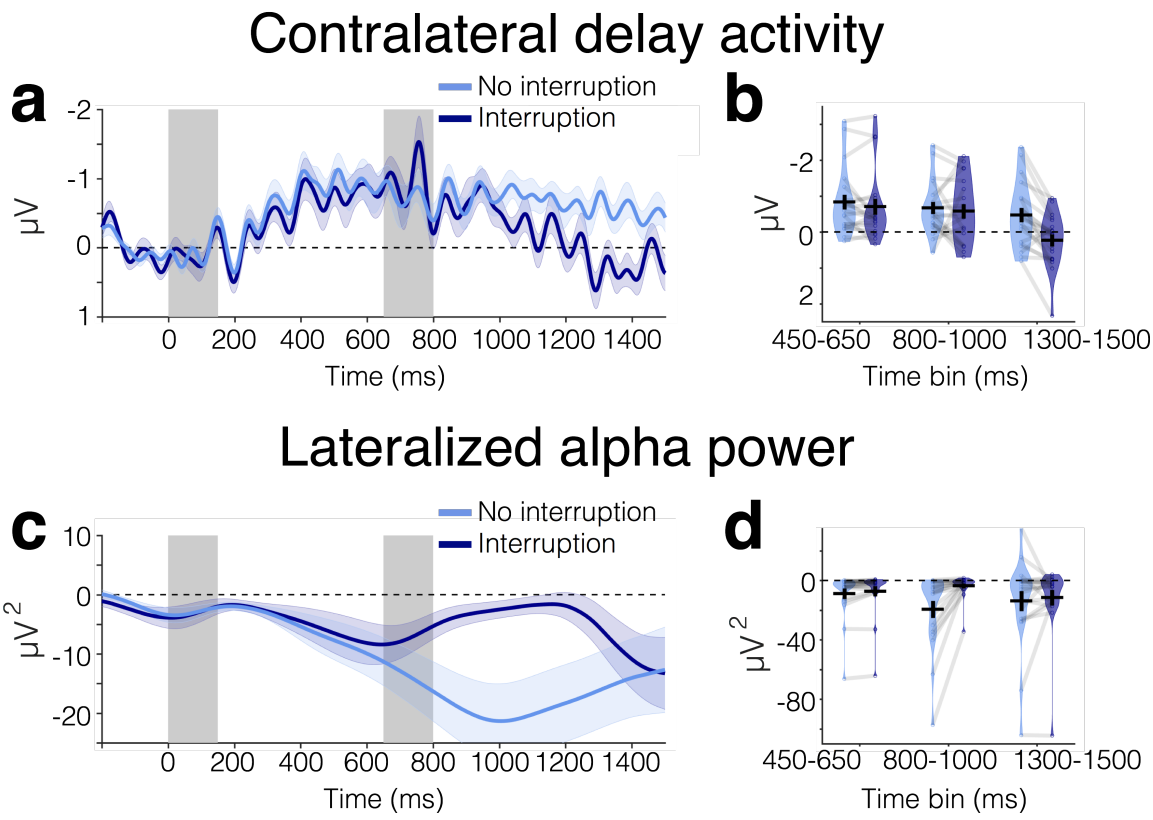

Figure 3-3. Contralateral delay activity and lateralized alpha power results for Experiment 3-1.

(a) CDA amplitude over time for trials with (dark blue line) and without (light blue line) interrupters. The light color envelopes around each line represent standard error of the mean for each condition. The first vertical gray bar (timepoint 0-150 ms) represents when the memory array was on the screen, and the second gray bar (timepoint 650-800 ms) represents when the interrupters were on the screen, if there were interrupters on that trial. (b) CDA amplitude for trials with (light blue plots) and without interruption (dark blue plots) averaged over the three time windows of interest (450-650 ms, 800-1000 ms, 1300-1500 ms). Each dot represents the mean number correct in one condition for one participant. A line connects data from the same participant. The horizontal black line depicts the mean across participants. The black error bars reflect the standard error of the mean. The shaded area reflects the density across participants. (c) Lateralized alpha power over time for trials with and without interruption (d) Average lateralized alpha power for trials with and without interruption.

Lateralized Alpha Power

Pre-Interruption (450-650 ms). Alpha power (Figure 3-3) was significantly more negative at contralateral compared to ipsilateral electrodes before interruption onset (450-650 ms) on trials with, $t(19)=-2.131$, $p=0.046$, $M=-7.264$, 95% CI (-14.398 -0.130), and without

interruption, t(19)=-2.517, p=0.021, M=-8.815, 95% CI (-16.145 -1.486). Alpha power was significantly more lateralized on trials without interruption than trials with interruption during this time window, t(19)= -2.573, p= 0.019, M= -1.551, 95% CI (-2.813 -0.289). We suspect that this may be due to time smearing in the alpha analysis. Time smearing is a side-effect of Fourier transforms, as the calculation of power at any time point incorporates data from time points before and after the time point of interest. Therefore, the effect of the interruption may be smeared across time, causing it to appear like there are differences in alpha power before interruption onset when there actually are only differences after interruption onset.

Post-Interruption I (800-1,000 ms). Immediately following the offset of the interruption (800-1,000 ms), lateralization of alpha power remained robust following trials without interruption, t(19)=-3.423, p=0.003, M= -19.393, 95% CI (-31.250 -7.536). However, alpha power was not significantly lateralized following trials with interruption, but this effect was trending towards significance, t(19)=-2.054, p=0.054, M=-3.554, 95% CI (-7.175 0.067). During this time window, lateralized alpha power was significantly more lateralized on trials without interruption than trials with interruption, t(19)= -3.629, p= 0.002, M=-15.839, 95% CI (-24.974 -6.704

Post-interruption II (1,300-1,500 ms). By the end of the trial (1,300-1,500 ms), however, there was no longer a significant difference in lateralized alpha power between the two conditions, t(19)= -0.904, p= 0.377, M= -3.879, 95% CI (-12.863 5.104). Lateralized alpha power was significantly lateralized in both conditions, No Interruption: t(19)=-2.144, p=0.045, M= -14.207, 95% CI (-28.077 -0.337), Interruption: t(19)=-2.137, p=0.046, M=-10.328, 95% CI (-20.444 -0.213).

Conclusions

In Experiment 3-1, participants' working memory performance was reduced when they were interrupted during the retention interval. In addition, both the CDA and lateralized alpha power were negatively impacted by the interrupters, but this effect had distinct time courses for the two signals. The CDA briefly sustained following interruption while alpha power immediately became less lateralized. By the end of the trial, CDA was no longer present, but alpha power re-lateralized. These results suggest that the CDA and lateralized alpha power respond distinctly to task-irrelevant interruptions.

Experiment 3-2

In Experiment 3-2, we sought to determine whether the neural responses to task-irrelevant interruptions could be modulated by task expectancy, or if they are fixed responses to interruption irrespective of the subject's expectations. Thus, in Experiment 3-2, we compared the same 25% interruption condition employed in Experiment 3-1 with one in which interrupters were presented on 75% of trials. We predicted that a higher frequency of task-irrelevant interruptions should allow participants to be better prepared for interruptions. Accordingly, CDA and lateralized alpha power should sustain for longer following interruption. Importantly, we will also examine the onset and offset of the CDA and lateralized alpha power to examine whether the time course of the two subprocesses may be affected differently.

Materials & Methods

Participants. Thirty novel volunteers, naïve to the objective of the experiment participated for payment (~15 USD per hour). All data were collected in a single session. The data of 10 participants were excluded from the analysis because of too many artifacts, poor

behavioral performance or technical problems (same criteria as in Experiment 3-1). The remaining 20 participants (11 male) were between the ages of 19-32 (M = 23.54, SD = 3.85).

Stimuli & Procedures. Stimuli were identical to Experiment 3-1 (Figure 3-1). Procedure was also identical to Experiment 3-1, except for the following changes. The retention interval was increased to 2000 ms. The experiment was divided in two halves in each of which the probability of interruption was varied. The order of the halves was counterbalanced across participants. The probability for interruption was 25% in one part (no interruption: 75%) and 75% in the other part (no interruption: 25%). This resulted in 2 × 2 design with the factors Interruption (no interruption versus interruption) and Probability (high versus low). Participants completed 1,920 trials in total (24 blocks of 80 trials each), 240 trials for each of the two low probability conditions and 720 trials for each of the two high probability conditions.

Analysis

Behavioral and EEG data were analyzed analogously to Experiment 3-1 but included the additional factor Probability. For statistical analyses, we forwarded the mean CDA amplitude (contra- minus ipsilateral activity) to a two-way ANOVA with the within-subjects factors Interruption (interruption versus no interruption) and Probability (75% probability for interruption versus 25%). Additionally, for the CDA and lateralized alpha analyses, the time window before probe onset was 1800-2000 ms, as we extended the retention interval by 500 ms.

Results

Behavior

Performance (Figure 3-4), as measured by K, was significantly worse on trials that were interrupted (Low Probability M = 1.4, High Probability M = 1.6) than on trials that were not interrupted (Low Probability M = 1.6, High Probability M = 1.7), regardless of probability,

significant main effect of Interruption, $F(1,19)=21.288$, $p<0.001$, $\eta_p^2=0.528$. There was not a significant main effect of Probability ($F(1,19)=3.575$, $p=0.074$, $\eta_p^2=0.158$) or interaction of Interruption and Probability ($F(1,19)=1.420$, $p=0.248$, $\eta_p^2=0.070$).

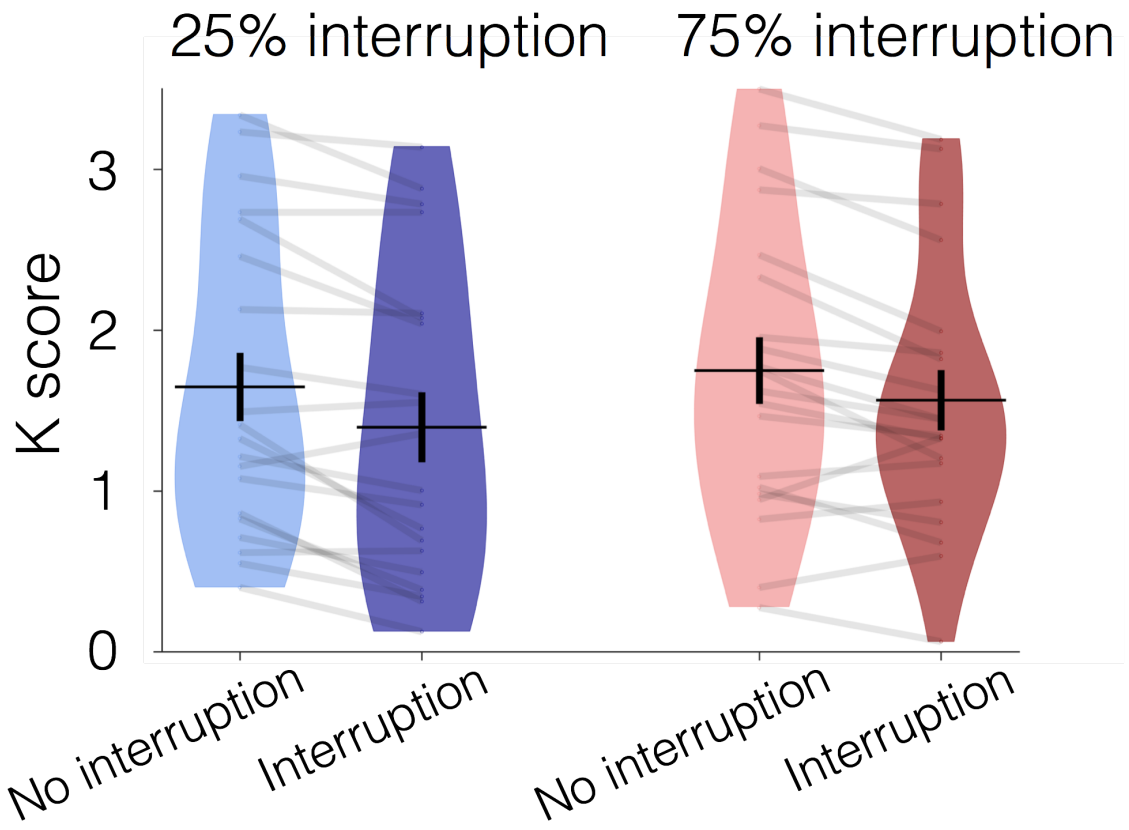

Figure 3-4. Behavioral results for Experiment 3-2.

Average performance is represented with the horizontal black line. The standard error is represented as the vertical black line. The distribution of K scores for all participants is represented by the violin plot. Dots and light gray lines represent one participant's performance.

Contralateral Delay Activity

Pre-interruption (450-650 ms). Before interruption onset (450-650 ms), there was a significant CDA in all four conditions, all one-sample t-tests $p\leq0.002$. Additionally, there was no difference in CDA amplitude between any of the conditions, $p\geq0.060$ for the main effects of

Interruption, Probability and their interaction. Though the interaction of Interruption and Probability was trending towards significance, $p=0.060$, $\eta_p^2=0.173$.

Post-interruption I (800-1,000 ms). Immediately following interruption offset (800-1,000 ms), the influence of interruption on CDA amplitude depended on the probability of being interrupted, significant interaction of Probability and Interruption: $F(1,19)=9.951$, $p=0.005$, $\eta_p^2=0.344$. Follow-up t-tests run separately for trials with and without interruption revealed that when interrupters were present, the amplitude of the CDA depended on the probability of interruption, $t(19)=2.252$, $p=0.036$. The CDA was significantly larger in the High Probability condition ($M = -0.660$) than in the Low Probability condition ($M = -0.265$). On trials without interruption, there was no difference in CDA amplitude between probabilities, $t(19)=-0.858$, $p=0.402$. The main effects of Interruption and Probability were not significant, $p\geq0.129$.

Post-interruption II (1,800-2,000 ms). By the end of the trial (1,800-2,000 ms), there was no difference in CDA amplitude between any of the conditions, $p\geq0.142$ for the main effects of Interruption, Probability, and their interaction. There was no longer a significant CDA in any condition, all one-way t-tests $p \geq 0.088$. CDA tends to decline over time, and by extending the delay period compared to Experiment 3-1, we may have reached the point at which the CDA tends to decline naturally (Vogel & Machizawa, 2004a).

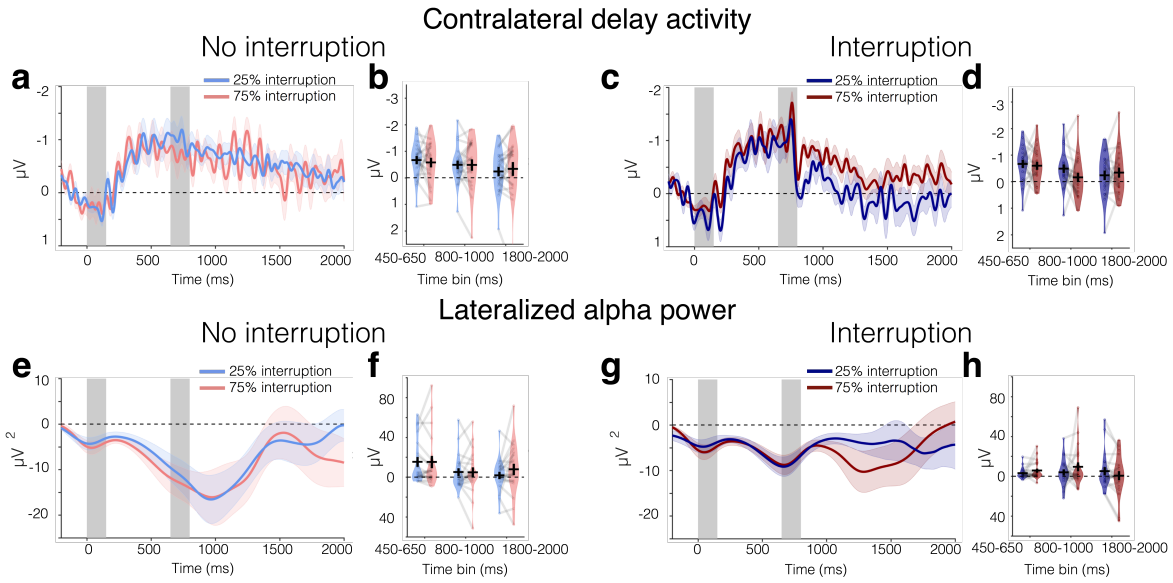

Figure 3-5. Contralateral delay activity and lateralized alpha power results for Experiment 3-2.

(a) CDA amplitude and (e) alpha power lateralization over time for trials without interruption in the 25% (light blue) and 75% (pink) interruption blocks. The light color envelopes around each line represent standard error of the mean for each condition. The first vertical gray bar (timepoints 0-150 ms) represents when the memory array was on the screen, and the second gray bar (timepoints 650-800 ms) represents when the interrupters were on the screen, if there were interrupters on that trial. (b) CDA amplitude and (f) alpha power lateralization for trials without interruption in the 25% (light blue plots) and 75% (pink plots) interruption blocks averaged over the three time windows of interest (450-650 ms, 800-1000 ms, 1800-2000 ms). Each dot represents the mean number correct in one condition for one participant. A line connects data from the same participant. The horizontal black line depicts the mean across participants. The black error bars reflect the standard error of the mean. The shaded area reflects the density across participants. (c) CDA amplitude and (g) lateralized alpha power over time for trials with interruption in the 25% (dark blue) and 75% (red) interruption blocks. (d) CDA amplitude and (h) alpha power lateralization averaged over three time windows of interest for trials with interruption in the 25% (dark blue plot) and 75% (red plot) interruption blocks.

Lateralized Alpha Power

Pre-interruption (450-650 ms). Alpha power (Figure 3-5) was significantly suppressed in all conditions before the interruption onset (450-650 ms), all one-sample t-tests $p \leq 0.005$. The influence of interruption on alpha power suppression depended on the probability of interruption, significant interaction of Probability and Interruption: $F(1,19)=4.881$, $p=0.040$, $\eta_p^2 = 0.204$. As in

Experiment 3-1, this pre-interruption difference could be due to time smearing of alpha power. There was no difference in lateralized alpha power between trials that were and were not interrupted for either the High ($f(19)=-2.046$, $p=0.055$) or Low ($f(19)=0.279$, $p=0.789$) probability trials, though this effect was trending in the High probability trials.

Post-interruption I (800-1,000 ms). Immediately following the offset of interruption (800-1,000 ms), alpha power was significantly lateralized in all conditions, all one-sample t-tests $p \leq 0.006$. There was a significant main effect of Interruption on the strength of alpha lateralization: $F(1,19)=6.530$, $p=0.019$, $\eta_p^2 = 0.256$. For both High and Low probability trials, lateralized alpha power was stronger on trials without interruption (Low Probability M= -15.498, High Probability M= -15.368) than on trials with interruption (Low Probability M= -4.549, High Probability M= -4.966). No other effects were significant, $p \geq 0.778$.

Post-interruption II (1,800-2,000 ms). By the end of the trial (1,800-2,000 ms), the influence of interruption on lateralized alpha power depended on the probability of interruption, significant interaction of Interruption and Probability, $F(1,19)=6.365$, $p=0.021$, $\eta_p^2 = 0.251$. The difference between lateralized alpha power in high and low probability trials is different with and without interruption. However, within High probability trials, follow-up t-tests revealed that there was not an alpha power suppression difference between trials that were interrupted and those that were not for either High ($t(19)=-1.923$, $p=0.070$) or Low ($t(19)=1.077$, $p=0.295$) probability trials. Also, alpha power suppression was not significantly different between High and Low probability trials for trials that were interrupted ($t(19)=-1.278$, $p=0.217$) or for those that were not interrupted ($t(19)=1.362$, $p=0.189$). The interaction of Interruption and Probability is disordinal because the interaction is significant, but the follow-up t-tests are not significant. Disordinal interactions indicate that a factor has one kind of effect in one condition and the

opposite kind of effect in the other condition. In this case, in the Low probability condition, alpha power is numerically more lateralized when interrupters were present (M=-5.14, SD=18.89) than when they were not present (M=-1.84, SD-16.53). However, in the High probability condition, alpha power is numerically less lateralized when interrupters were present (M=-0.58, SD=20.83) than when they were not present (M=-7.296, SD=25.92). The main effects of Interruption and Probability were also not significant, $p \geq 0.472$.

Conclusions

In Experiment 3-2, we found that behavioral performance was worse when participants were interrupted than when they were not interrupted regardless of the probability of interruption. Once again, the neural results revealed that both CDA and lateralized alpha power were negatively impacted by interruptions, but these two signals had distinct time courses. Following interruption, CDA sustained, but lateralized alpha power became less lateralized. Additionally, the amplitude of CDA immediately following interruption depended on the probability of interruption – CDA amplitude was larger when participants were expecting to be interrupted. However, alpha power lateralization did not depend on expectations – alpha power shifted towards baseline when interruptions were present, regardless of the probability of interruption.

By the end of the trial, CDA was no longer present on trials with interruptions, regardless of probability. This replicates the CDA results from Experiment 3-1. The effect of probability and interruption on alpha power lateralization by the end of the trial were a bit more ambiguous. In the Low probability block, lateralized alpha power was equivalent on trials with and without interruptions. This replicates the results from Experiment 3-1. However, upon visual inspection

of the results, the "recovery" pattern following interruption was not as apparent. This is because overall alpha power lateralization on trials without interruption was very close to baseline, unlike in Experiment 3-1 where alpha power was robustly lateralized. This reduction in alpha power lateralization on trials without interruption could plausibly be due to the length of the retention interval. In Experiment 3-1, the retention interval was 1,500 ms, and in Experiment 3-2 it was extended to 2,000 ms. We did this so that we could investigate whether CDA would return if participants had more time post-interruption. However, both CDA and alpha power lateralization tend to shift towards baseline with longer delays, which may be the reason why alpha power is less lateralized by the end of the trial in Experiment 3-2 than in Experiment 3-1. Regardless of the amount of alpha power lateralization at the end of the trial, we still found a significant reversal of the effect of probability and interruption in the High probability condition as compared to the Low probability condition. In the High probability condition, alpha power was more lateralized on trials without interruptions than on trials with interruptions, but this effect was reversed in the Low probability condition.

Discussion

Working memory maintains information so that it can be used despite momentary perturbations from task-irrelevant information. Here, we examined how memory representations that have already reached a stable state respond to visual interruption. As expected, we found a modest behavioral impact of interruption. Participants remembered significantly fewer items when they were interrupted than when they were not interrupted, but they performed above chance in all conditions. Despite a modest behavioral impact, task-irrelevant interruption produced substantial perturbations on two well-characterized EEG signals of working memory,

lateralized alpha power and contralateral delay activity (CDA). Both lateralized alpha power, an index of sustained spatial attention, and CDA, an index of actively maintained working memory representations, were disrupted at certain points during the delay, but the time course of these perturbations varied. Lateralized alpha power results suggest that attention shifted towards baseline immediately following the interruption but had returned to the target positions by the end of the trial. By contrast, the CDA results suggest that working memory representations continued to persist following the interruption but was eliminated by the end of the trial. We additionally found that task expectancy modulated the timing and magnitude of these perturbations of working memory representations, suggesting that the brain's response to task-irrelevant interruption is regulated by task context. The distinct time courses of and the influence of task context on lateralized alpha power and CDA have many interesting theoretical implications that future work can help elucidate.

Neural response immediately following interruption

Sudden onsets of task-irrelevant interruption have been shown to capture attention when interrupters are visually salient (Andrews et al., 2009; Bisley & Goldberg, 2003; van Moorselaar et al., 2018). In our experiment, we used lateralized alpha power as an index of sustained spatial attention (Foster et al., 2016; Hakim et al., 2018). Following the onset of interruption, lateralized alpha power almost immediately shifted towards baseline. When lateralized alpha power is at baseline, it suggests that participants are no longer spatially attending the lateral memory items. Neural evidence from previous studies suggests that participants attend to the locations of interrupting stimuli (Bisley & Goldberg, 2003; van Moorselaar et al., 2018) because of attentional capture (Feldmann-Wüstefeld & Schubö, 2013; Sawaki & Luck, 2012). Thus, in the

present study, participants presumably shifted their attention away from lateralized representations following the onset of task-irrelevant interruption to the centrally presented interrupters.

During this same time window, CDA remained robust and significantly above baseline. Previous research has shown that CDA is sensitive to trial-by-trial fluctuations in working memory performance and tracks the number of maintained object representations (Adam et al., 2015a; Ikkai et al., 2010). Considering this, the robust CDA immediately following the onset of interruption suggests that object representations persist, at least momentarily, following the withdrawal of spatial attention to a new position. The presence of CDA and lack of lateralized alpha power immediately following interruption raises the long-standing theoretical question of whether object representations maintained in working memory can persist without sustained spatial attention. Previous research has suggested that spatial attention is a rehearsal mechanism that facilitates the maintenance of object representations held in working memory (Williams & Woodman, 2012). Additionally, the positions of object representations are maintained in working memory even when spatial information is completely irrelevant (Foster, Bsales, et al., 2017a). Together, these previous results suggest that spatial attention aids the maintenance of working memory information, but do not address whether working memory representations necessitate sustained spatial attention. In the present study, the robust CDA and lack of lateralized alpha power following the onset of interruption suggest that object representations maintained in working memory can persist without sustained spatial attention. Therefore, our results suggest that working memory representations may not necessitate sustained spatial attention. Nevertheless, working memory representations may still be volatile without sustained spatial attention, given that CDA goes to baseline by the end of trials with interruption.

Neural activity at the end of interrupted trials

In this study, we sought to interrupt participants after working memories reached a stable state. Therefore, it is not surprising that participants can still perform well above chance in the interrupter-present trials. It is likely that interruptions to the WM representations at earlier moments, such as before CDA is fully formed, would produce larger behavioral decrements. Nevertheless, by the end of interrupted trials, we observed that the CDA was no longer reliable, but alpha power became re-lateralized. There is a large body of research that has shown that CDA tracks the active maintenance of information, is sensitive to trial-by-trial fluctuations in working memory performance, and distinguishes stable individual differences in working memory (Vogel and Machizawa, 2004; Luria *et al*., 2016). Therefore, the pattern of activity at the end of the trial suggests that participants re-oriented their attention to the locations of the memoranda, but no longer maintained active working memory representations. If participants no longer maintained object representations that are tracked by CDA, how were they able to still perform the change detection task on interrupted trials (albeit worse than non-interrupted trials)? There are a few possible explanations.

One possible explanation for the absence of the CDA at the end of the trial but above chance behavioral performance is that performance on interrupted trials could rely on offline memory representations. Previous research has shown that information in working memory can be simultaneously maintained in both active and passive memory states (Mallett & Lewis-Peacock, 2018). Therefore, when actively maintained memory traces are no longer present, information could still be retrieved from an offline state. Research that has investigated retrieval of information from offline memory states has found that alpha power tracks information

retrieved from long-term memory (Fukuda et al., 2016). Additionally, other research has suggested that attention can aid recall of information that would be otherwise unavailable to working memory (Murray et al., 2013). These findings dovetail with our results – at the end of interrupted trials, when information about the memoranda is required to respond to the probe, lateralized alpha power could be re-instated in order to re-load information from offline memory storage, thereby bolstering behavioral performance. An alternate explanation for the recovery of lateralized alpha power at the end of the trial is that it reflects the anticipation of the upcoming memory probe. The memory probe always appeared in the same location as one of the memory items. Thus, in order to shift attention to the location of the upcoming probe, participants had to remember the locations of the original memory items. Therefore, even if the re-lateralization of alpha power at the end of interrupted trials reflects the orienting of spatial attention to the location of the anticipated memory probe, it still suggests that this re-lateralization relies on the retrieval of task-relevant spatial information. Both the re-loading and re-orienting explanations of the recovery of alpha power are plausible and theoretically interesting explanations for this pattern of activity.

The relatively good behavioral performance without CDA could alternately be explained by other neural traces of actively maintained working memory representations that we are not measuring. The CDA is a coarse neural measure that compares activity contralateral and ipsilateral to memory items. Thus, it is not an exhaustive measure of working memory. More spatially global neural signals or more distributed patterns of activity, for example, could sustain following task-irrelevant interruption, and these signals could plausibly bolster behavioral performance. Regardless of the mechanism that preserves information about the memoranda, our

results strongly suggest that actively maintained information is dynamically perturbed following task-irrelevant interruption.

Modulation of CDA and alpha power by task demands

In Experiment 3-2, we varied task demands by interrupting participants on 75% (high) or 25% (low) of trials. Following interruption onset, we found the same pattern of result as Experiment 3-1; Lateralized alpha power shifted towards baseline while CDA persisted. However, the amplitude of the CDA varied as a function of task demands. When task demands were high, CDA amplitude was higher than when task demands were low. This suggests that participants were able to better protect working memory representations when they were expecting to be interrupted, and that task context is involved in how the brain responds to task-irrelevant interruption. On the other hand, the influence of task demands on lateralization of alpha power was more ambiguous. Our results suggest that spatial attention may be uniformly captured by interrupters initially regardless of expectation. However, during certain points in the trial, lateralization of alpha power may vary as a function of task demands. Therefore, the neural responses to interruption that we observed were affected both by both interruption and task demands. These results go hand-in-hand with previous research that has shown that interrupter by salient irrelevant stimuli can be modulated by top-down control. For example, when a color singleton is presented on 20% of the trials, it slows down response times in a visual search task more than when it is presented on 50% of trials (Folk & Remington, 2015; Horstmann, 2005; Marini et al., 2013; Müller et al., 2009) because attention requires more time to be deployed to the relevant information when rare interrupters appear (Töllner et al., 2012).

Conclusions

In this set of experiments, we investigated the impact of task-irrelevant interruption on two dissociable neural signals, CDA, a neural index of actively maintained representations, and lateralized alpha power, an index of sustained spatial attention. By tracking these neural markers of working memory, we were able to observe changes in active representations that would not be apparent from behavioral measures alone. Both CDA and lateralized alpha power were impacted by task-irrelevant information yet had distinct time courses. Our results suggest that following interruption, lateralized visual representations of memoranda can stay active in working memory for a short period of time without lateralized spatial attention before they are lost. These representations do not recover by the end of the trial and are presumed to be stored offline. By contrast, attention is directed away from the spatial location of memoranda immediately after the onset of the interruption but can recover later and may even contribute to the retrieval of information from offline storage. Thus, our results show that task-irrelevant interruption could motivate the transfer of information from active to passive storage. Moreover, the dissociation between CDA and lateralized alpha power further emphasizes that these neural markers distinctly contribute to the maintenance of information in working memory and may distinctly protect actively maintained memories from interruption.

CHAPTER 4.

Relevant and irrelevant interruptions distinctly impact spatial attention and object-based storage

A critical feature of working memory is to protect internal representations from external interference. For example, when driving, the working memory representation of our route must be maintained despite irrelevant external interference, such as a flashing colorful billboard. Nevertheless, external information is sometimes relevant, and working memory must integrate this new information with our ongoing working memory representations. For example, a flashing sign warning us of a car accident ahead may capture our attention – but to our advantage. This new information allows us to update our working memory representation of our route in order to avoid traffic caused by the car accident. Attentional control mechanisms help determine which information gets access to our capacity-limited working memory system.

Models of attentional control suggest that attentional selection is based on a competitive process (Bundesen, 1990; Duncan & Humphreys, 1989) in which both goal-driven and stimulus-driven factors determine which information is selected from a given display (Awh et al., 2012; Corbetta & Shulman, 2002; Desimone & Duncan, 1995; Egeth & Yantis, 1997). But can goal-driven attentional selection override stimulus-driven capture? Some previous research has suggested that salient irrelevant interrupters capture attention before it can be suppressed (Feldmann-Wüstefeld & Schubö, 2013; Liesefeld et al., 2017; R. Sawaki et al., 2012). Based on this, they argue that attentional capture is obligatory because salient information needs to be processed before being discarded. Other research suggests that interrupter suppression can prevent attentional capture (Feldmann-Wüstefeld et al., 2020; Gaspar & McDonald, 2014;

Gaspelin et al., 2015, 2017; Gaspelin & Luck, 2018). This work has found that, for example, participants make fewer erroneous eye movements towards (Gaspelin et al., 2017) and are less likely to report the identity of (Gaspelin et al., 2015b) a successfully suppressed salient interrupter than a less salient interrupter that was not suppressed. Although these appear to be conflicting perspectives, we suggest here that they might be reconciled by distinguishing between two distinct forms of attentional capture.

Attentional capture and suppression are often treated as a monolithic process: the onset of a stimulus results in either capture or suppression of the novel information. However, recent work suggests that attention may include two distinct sub-component processes: attention to regions in space and representations of objects that occupy the attended regions (Hakim, Adam, et al., 2019; Maxwell et al., 2020; Prinzmetal et al., 2009; Zivony & Lamy, 2018). In line with these studies, we propose that involuntary attentional capture may also be broken down into at least two distinct sub-component processes: 1) Spatial capture, which refers to when spatial attention shifts towards the location of irrelevant stimuli 2) Item-based capture, which refers to when item-based representations of interrupting stimuli are formed in working memory. In the current work, we obtained EEG evidence to directly test whether spatial capture is distinct from item-based capture.

We separately measured spatial and item-based capture using EEG markers of spatial and item-based attentional capture. We used lateralized alpha power (8-12 Hz) to track spatial capture (Foster et al., 2016; Hakim, et al., 2019). This oscillatory signal has been shown to track attended hemifield (Hakim, et al., 2019) and has been shown to contain precise spatial information about attended stimuli (Foster et al., 2016, 2017). To track item-based capture, we used the contralateral delay activity (CDA) and the distractor positivity (Pd). The CDA tracks the

number of items maintained in working memory (Balaban & Luria, 2017b; Luria et al., 2016b), whereas the Pd tracks the suppression of irrelevant information (Burra & Kerzel, 2014; Feldmann-Wüstefeld & Schubö, 2013; Hickey et al., 2009), while also being sensitive to the number of irrelevant items that are presented (Feldmann-Wüstefeld et al., 2019).

We used these EEG signals to assess how salient items with a sudden onset (interrupter) are processed when subjects are maintaining relevant information in working memory. In addition, we manipulated the task-relevance of the new, interrupting information to compare how each type of attentional capture is influenced by goal-driven selection. Finally, while past work has typically focused on competition between simultaneously presented targets and interrupters, here we focused on how task-relevant and task-irrelevant interrupters influence the maintenance of items that have already been stably encoded into working memory. This provided the opportunity to obtain clear evidence regarding the degree to which interrupters elicited spatial and item-based attentional capture, and the distinct impact of task-relevance on each form of capture.

Prominent models of attentional control assert that visually selected stimuli should automatically gain access to working memory at least for a short period of time (e.g., Bundesen et al., 2005). Moreover, previous research has shown that a sudden onset of salient but irrelevant information captures attention (Feldmann-Wüstefeld et al., 2015; Franconeri & Simons, 2003; Theeuwes, 2010; Yantis & Jonides, 1984) and that this negatively impacts ongoing working memory representations (Bisley & Goldberg, 2010; Hakim, Feldmann-Wüstefeld, et al., 2019; van Moorselaar et al., 2017b). Does this interruption of working memory maintenance reflect an obligatory encoding of this new information into working memory? To anticipate the results, we observed clearly distinct effects of task relevance on spatial and item-based attentional capture.

Continuous tracking of alpha laterality showed that spatial attention was captured by interrupting stimuli, regardless of whether they were relevant or not. In sharp contrast, item-based attentional capture was completely determined by task relevance. Task-relevant interrupters were encoded into working memory, as shown by N2pc and CDA signals that tracked item individuation and working memory maintenance, respectively. By contrast, a P_D was observed contralateral to the task-irrelevant interrupters and no CDA was observed. Thus, task-irrelevant interrupters were not actively encoded into working memory, even though they clearly captured spatial attention. Thus, our findings offer a potential reconciliation of prior conflicting findings by showing that the encoding of interrupting information into working memory can be suppressed even when there is clear evidence that spatial attention has been captured.

Materials & methods

Experiment 4-1

In Experiment 4-1, we sought to determine how task-relevant versus task-irrelevant interrupters are processed. To this end, we presented memory array items along the midline and presented interrupters laterally. This allowed us to isolate the neural representations of the interrupters themselves. With this design, any lateralized signal, such as CDA or lateralized alpha power, should reflect the processing of the interrupters and not the memory array. Previous research has shown that active representations may be required for the identification of relevant stimulus features (Mazza et al., 2007; Mcdonald et al., 2013). Therefore, we predicted that when participants had to discriminate the task-relevant interrupters (discriminate condition), they would be more likely to encode them into working memory than when task-irrelevant interrupters were presented (ignore condition). Accordingly, there should be a CDA following task-relevant, but not following task-irrelevant interrupters. Conversely, when participants could

ignore the task-irrelevant interrupters, we predicted that they would actively suppress them, as their features do not need to be identified. The distractor positivity (P_D) has been shown to track suppression of irrelevant information (Burra & Kerzel, 2014; Feldmann-Wüstefeld & Schubö, 2013; Sawaki & Luck, 2012). Therefore, we should expect to find a robust P_D when participants ignore the task-irrelevant interrupters, but not when they discriminate the task-relevant interrupters.

Participants. Thirty novel volunteers, naïve to the objective of the experiment participated for payment ($15 USD per hour). Data from one participant was excluded from the analysis because of technical issue with the behavioral data file. Data from nine participants were excluded from the analysis because of too many artifacts that resulted in fewer than 150 trials in any condition. The remaining 20 participants (6 male) were between the ages of 21-31 (M = 23.5, SD = 3.3). Participants in all experiments reported normal or corrected-to-normal visual acuity as well as normal color vision. All experiments were conducted with the written understanding and consent of each participant. The University of Chicago Institutional Review Board approved experimental procedures.

Stimuli. All stimuli were presented on a gray background (~33.3 cd/m^2). Cue displays showed a central fixation dot (0.2° × 0.2°). Memory displays showed four colored squares (1.1° by 1.1°, mean luminance 43.1 cd/m^2) along the midline with a randomly jittered horizontal offset of maximally 0.55° (half of an object). Colors for the squares were selected randomly from a set of 11 possible colors (Red = 255, 0, 0; Green = 0, 255, 0; Blue = 0, 0, 255; Yellow = 255, 255, 0; Magenta = 255, 0, 255; Cyan = 0, 255, 255; Purple = 102, 0, 102; Brown = 102, 51, 0; Orange =

255, 128, 0; White = 255, 255, 255; Black = 0, 0, 0). No color was repeated. On 50% of trials, the retention interval display remained blank with a central fixation dot (0.2° × 0.2°). However, on the other 50% of trials, during the delay the interrupting stimuli appeared laterally. On one side of the screen, four colored circles (25% of all trials) or four squares (25% of all trials) appeared during the delay. Items from these colors were chosen from the say 11 possible colors but were never the same as the memory array items on a given trial. Additionally, these items had the same area as the items from the memory array. On the other side of the screen, four gray diamonds appeared (RGB= 80, 80, 80) at the same time as the colored circles/squares. These gray diamonds were the same area as the colored circles/ squares. The gray diamonds were also luminance matched (i.e. iso-luminant) to the average of the colored circles/squares, so as to achieve a comparable bottom-up saliency. We presented these gray diamonds so as to match the bottom-up visual stimulation on both sides of the screen. The hemifield in which the diamonds and the hemifield in which the colored circles/squares were presented were randomly selected each trial. All stimuli had the same area. Probe displays showed one colored square along the midline in the same location as one of the memory array items, randomly picked, in the original array. In 50% of the trials, the color of the square in the attended hemifield was identical (no change trial) to the memory display. In the remaining 50% of trials, it was one of the colors not used in the memory or interrupter display (change trials).

Apparatus. Participants were seated with a chinrest in a comfortable chair in a dimly lit, electrically shielded and sound attenuated chamber. Participants responded with button presses on a standard keyboard that was placed in front of them. Stimuli were presented on an LCD computer screen (BenQ XL2430T; 120 Hz refresh rate; 61 cm screen size in diameter; 1920 ×

1080 pixels) placed at 74 cm distance from participants. An IBM-compatible computer (Dell Optiplex 9020) controlled stimulus presentation and response collection.

Procedure. Each trial began a memory display consisting of four colored squares along the midline appeared for 150 ms. Participants were instructed to memorize as many colored squares in the memory display as possible. Participants had to remember the items over a blank retention interval that contained a central fixation dot. The retention interval lasted 2,000 ms, regardless of whether interrupters appeared. In 50% of the trials, a interrupter display appeared 500 ms after memory display offset for 150 ms. The interrupting display consisted of four circles (50% of interrupter trials) or four squares (50% of interrupter trials) that appeared laterally. This display was visually balanced by four iso-luminant gray diamonds that appeared in the opposite hemifield. In half of the trials, participants were instructed to ignore the interrupter displays (Ignore block). In the other half of the trials, participants were instructed to discriminate the shape of the interrupter items (Discriminate block). After the retention interval, a probe display appeared until response. In both Ignore and Discriminate blocks, participants had to indicate whether the object at the probed location of the attended hemifield changed color ("?/" key) or did not change color ("z" key). In the Discriminate block, participants additionally performed a go-no-go task. Before responding to the probe, they had to press "space" to indicate that the interrupter objects were circles. If the interrupter objects were squares, they did not have to press any key. After participants responded, the next trial started after a blank inter-trial interval of 750 ms. Participants completed a total of 1600 trials (20 blocks of 80 trials), i.e., 800 trials with interrupters and 800 trials without interrupters. The first half of the experiment was always the Ignore blocks (task-irrelevant interrupters), and the second half of the experiment was always the

Discriminate blocks (task-relevant interrupters). Information about average performance and a minimum break of 30 seconds was provided after each block. See Figure 4-1 for a visual depiction of the task.

We presented the interrupters in locations that did not overlap with the locations of the memory items to avoid visual masking. Importantly, the relative position of interrupters and targets matter in this kind of change detection tasks. When interrupters are presented laterally with targets on the vertical midline, the neural signature of sustained distractor suppression can be isolated (CDAp). Conversely, when interrupters are presented on the vertical midline and targets are presented laterally, the neural signature of target processing can be isolated (Feldmann-Wüstefeld & Vogel, 2018). Accordingly, in Experiment 4-1, we were interested in the neural representations of the interrupters, so we placed the interrupters laterally. Thus, lateralized signals, such as CDA and lateralized alpha power, could be used to assess item-based and spatial capture elicited by the lateralized interrupters.

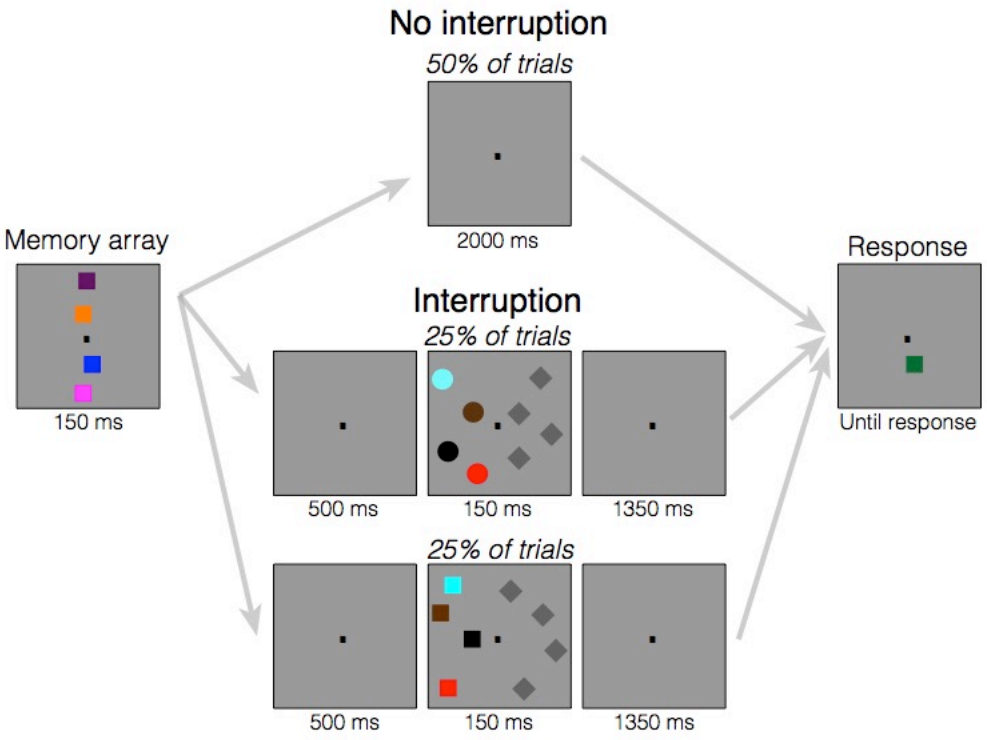

Figure 4-1. Task design for Experiment 4-1.

At the start of each trial, the memory array appeared, which consisted of four colored squares along the midline. Participants were told to remember the colors and locations of these squares over the brief delay. Following the memory array, the screen went blank. Then, either the screen remained blank the entire delay (no interrupter condition) or the screen remained blank for 500 ms (interrupter condition) followed by a series of interrupters presented laterally. This interrupter array consisted of either four colored squares or circles on one side of the screen and iso-luminant gray diamonds on the other side of the screen. When participants were in the "Ignore" condition, they were told to always ignore these task-irrelevant interrupter objects. When they were in the "Discriminate" condition, they were told to determine the shape of the colored stimuli (squares vs. circles) in order to report whether the stimuli were circles. They were told to withhold their response until the response screen appeared. Following the interrupter array, the screen then went blank for the rest of the delay. On the final screen, one square on the midline re-appeared and could either be the same color as the original square or it could be a different color. In both conditions, participants had to report whether the square on the cued side of the screen changed colors. In the "Discriminate" condition, participants additionally had to report whether the interrupter were circles, if there were interrupters on that trial.

Artifact rejection.

We recorded EEG activity from 30 active Ag/AgCl electrodes (Brain Products actiCHamp, Munich, Germany) mounted in an elastic cap positioned according to the

International 10-20 system [Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz]. FPz served as the ground electrode and all electrodes were referenced online to TP10 and re-referenced off-line to the average of all electrodes. Incoming data were filtered [low cut-off = .01 Hz, high cut-off = 80 Hz, slope from low- to high-cutoff = 12 dB/octave] and recorded with a 500 Hz sampling rate. Impedances were kept below 10kΩ. To identify trials that were contaminated with eye movements and blinks, we used electrooculogram (EOG) activity and eye tracking. We collected EOG data with 5 passive Ag/AgCl electrodes (2 vertical EOG electrodes placed above and below the right eye, 2 horizontal EOG electrodes placed ~1 cm from the outer canthi, and 1 ground electrode placed on the left cheek). We collected eye-tracking data using a desk-mounted EyeLink 1000 Plus eye-tracking camera (SR Research Ltd., Ontario, Canada) sampling at 1,000 Hz. Usable eye-tracking data were acquired for 20 out of 22 participants in Experiment 4-1 and 29 out of 30 participants in Experiment 4-2.

EEG was segmented offline with 2000 ms segments time-locked to memory display onset, including a 200 ms pre-stimulus baseline period. Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic detection criteria to each segment. After automatic detection (see below), trials were manually inspected to confirm that detection thresholds were working as expected. Incorrect trials and any trial contaminated with artifacts were excluded from analysis. For example, a trial including a task-relevant interrupter was only considered as "correct" and included in further analyses if both the response to the primary and the interrupting task were correct. The removal of all trials with artifact(s) allowed us to ensure that there was not any missing data within any of the included trial epochs.

For the participants used in analyses, we rejected on average 12.0% (SD = 12.7%) of trials in Experiment 4-1, and 27.0% (SD = 6.3%) of trials in Experiment 4-2 per person. Collapsed across conditions, this resulted in an average of 1408 remaining trials per person in Experiment 4-1 and 1168 remaining trials in Experiment 4-2. In order to achieve an acceptable signal-to-noise ratio, participants were excluded if fewer than 150 correct trials were available in any of the conditions. Collapsed across conditions, participants in Experiment 4-1 were correct on 79.9% of all trials and in Experiment 4-2, they were correct on 78.7% of trials. Broken down by condition, participants in Experiment 4-1 were correct on 82.4% of trials without interrupters and 77.3% of interrupter-present trials. In Experiment 4-2, they were correct on 83.2% of trials without interrupters and 74.3% of interrupter-present trials.

Eye movements. We used a sliding window step-function to check for eye movements in the HEOG and the eye-tracking gaze coordinates. For HEOG rejection, we used a split half sliding window approach. We slid a 100 ms time window in steps of 10 ms from the beginning to the end of the trial. If the change in voltage from the first half to the second half of the window was greater than 20 μV, it was marked as an eye movement and rejected. For eye-tracking rejection, we applied a sliding window analysis to the x-gaze coordinates and y-gaze coordinates (window size = 100 ms, step size = 10 ms, threshold = 0.5° of visual angle).

Blinks. We used a sliding window step function to check for blinks in the VEOG (window size = 80 ms, step size = 10 ms, threshold = 30 μV). We checked the eye-tracking data for trial segments with missing data-points (no position data is recorded when the eye is closed).

Drift, muscle artifacts, and blocking. We checked for drift (e.g., skin potentials) by comparing the absolute change in voltage from the first quarter of the trial to the last quarter of the trial. If the change in voltage exceeded 100 μV, the trial was rejected for drift. In addition to slow drift, we checked for sudden step-like changes in voltage with a sliding window (window size = 100 ms, step size = 10 ms, threshold = 100 μV). We excluded trials for muscle artifacts if any electrode had peak-to-peak amplitude greater than 200 μV within a 15 ms time window. We excluded trials for blocking if any electrode had at least 30 time-points in any given 200-ms time window that were within 1V of each other.

Behavioral data analysis.

We separately analyzed performance for four separate conditions: trials without interrupters in the ignore block, trials with interrupters in the ignore block (task-irrelevant interrupters), trials without interrupters in the discriminate block, and trials with interrupters (task-relevant interrupters) in the discriminate block. Performance was converted to a capacity score, K, calculated as N x (H-FA), where N is the set-size, H is the hit rate, and FA is the false alarm rate (Cowan, 2001). To compare performance between conditions, we used a two-way ANOVA with the within-subjects factors Interruption (relevant versus irrelevant) and Relevance (ignore versus discriminate). All analyses were done with circle and square interrupters collapsed, since the circles and squares were equiprobable (circles: 50% of interrupter trials, squares: 50% of interrupter trials). We additionally ran two-tailed follow-up t-tests when it was justified.

Lateralized ERP analyses

Segmented EEG data was baselined from 200 ms to 0 ms before the onset of the memory displays. Artifact-free EEG segments were averaged separately for the two conditions when interrupters appeared (irrelevant versus relevant). Data was not analyzed for trials without interrupters because "laterality" was undefined in this condition. The difference between contralateral and ipsilateral activity for the electrode pair PO7/PO8 was calculated (i.e., the CDA), resulting in two average waveforms for each participant (one per analyzed condition). The average CDA amplitude was calculated for three time windows: before interrupter onset (400-650 ms), and two windows following interrupter offset (850-950 ms and 1050-1300ms). Previous research has shown that CDA amplitude should stabilize approximately 400 ms after memory array onset. To measure CDA after it stabilized, we chose a time window starting 400 ms after onset of the memory array. We wanted to measure an analogous time window following the onset of the interrupters (i.e., 400 ms after interrupter onset), which is why we chose the time window 1050-1300ms. For all of these time windows, we then compared the CDA across conditions with a paired-samples t-test. To measure the robustness of the CDA for each condition (reliable difference between contra- and ipsilateral activity), we also ran t-tests (against zero) for each time window and condition. These t-tests are two-tailed, unless otherwise stated. We corrected for multiple comparisons using a Bonferroni correction. We applied this correction to the two post- interrupter time windows. Thus, p-values less than 0.025 from these two post-interrupter time windows are considered significant. The pre- interrupter time window (400-650 ms) was not included in our correction for multiple comparison because it is logically impossible for there to be differences across conditions in this time window because participants were not able to predict when they would be interrupted on any given trial. We analyzed this time window so as to obtain a measure of noise. We would also like to note that a Bonferroni correction in this

case is extremely conservative, as we had strong *a-priori* expectations about when to expect an effect in the post- interruption I (850-950 ms) time window (Feldmann-Wüstefeld & Vogel, 2018). We also chose the post- interruption II (1050-1300ms) time window before analyzing the data, though this analysis window was more exploratory.

On trials with interrupters, we additionally analyzed the distractor positivity (P_D) and the N2pc. To calculate these signals, we used a data-driven approach from previous research to specify the specific time windows of interest (Feldmann-Wüstefeld & Vogel, 2018). To calculate the lateralized waveform (contra- minus ipsilateral to colored interrupters) for electrodes PO7/PO8, across participants and conditions. We determined the peak of the P_D and N2pc as the most positive or negative peak, respectively, 200 to 350 ms after interrupter onset across both conditions. The average amplitude from 20 ms before to 20 ms after that peak was used for statistical analyses on the P_D and the average amplitude from 50 ms before to 50 ms after that peak was used for statistical analyses on the N2pc.

Lateralized alpha power analysis

For the alpha power analyses, we did not baseline the segments. The raw EEG signal was band-pass filtered in the alpha band (8-12 Hz) using a two-way least-squares finite-impulse-response filter ("eegfilt.m" from EEGLAB Toolbox). Instantaneous power was then extracted by applying a Hilbert transform ('hilbert.m') to the filtered data. The resulting data were averaged separately for the two conditions when interrupters appeared (relevant versus irrelevant interrupters) and each laterality (contra- versus ipsi-lateral to cued hemifield) for the electrode pair PO7/PO8. Average alpha power was calculated for two of the same time windows as the CDA analysis: before interrupter onset (400-650 ms), and post- interrupter offset I (850-950 ms).

Previous research has not investigated lateralized alpha power while participants maintained working memory representations that were presented centrally. Therefore, even though we chose our time windows before analyzing our results, in this experiment, we did not have strong a-priori predictions about the timing of alpha power lateralization following the onset of lateralized interrupters. Additionally, even when relevant memoranda are presented laterally, alpha power takes up to 1,000 ms after memory array onset to become fully lateralized (Hakim, Adam, et al., 2019; Hakim, Feldmann-Wüstefeld, et al., 2019). Therefore, the third time window that we analyzed extended to included up to 1,000 ms after interrupter onset (1050-1650 ms). We then compared alpha power lateralization for each time window with a paired-samples t-test. To measure the robustness of alpha power lateralization for each condition (reliable difference between contra- and ipsilateral activity), we also ran t-tests (against zero) for each time window and condition. These t-tests are two-tailed, unless otherwise stated. Once again, we corrected the two post- interrupter time windows for multiple comparisons using a Bonferroni correction for two comparisons (significance threshold: $p<0.025$).

Results

Behavior

Behavioral performance was significantly above chance in all conditions (all one-sample t-tests, $p\leq0.001$). Participants remembered fewer items when they were interrupted ($M=1.95$, $sd=0.61$) than when they were not interrupted ($M=2.66$, $sd=0.60$). This difference was larger when the interrupters had to be discriminated ($\Delta M=1.001$, $sd=0.297$) than when they could be ignored ($\Delta M=0.419$, $sd=0.293$). This was evident from the significant interaction of Interruption and Relevance ($F(1,19)=55.725$, $p<0.001$, $\eta_p^2=0.746$), and from the significant follow-up paired-samples t-test ($t(19)=-7.465$, $p<0.001$). This t-test compared the difference between trials with

and without interrupters in the Ignore and the Discriminate conditions. The main effects of Interruption and Relevance were also significant (both $p<0.001$). Additionally, participants were 95.0% (±1.4%) accurate on the interrupter task. Behavioral results for Experiment 4-1 are depicted in Figure 4-2.

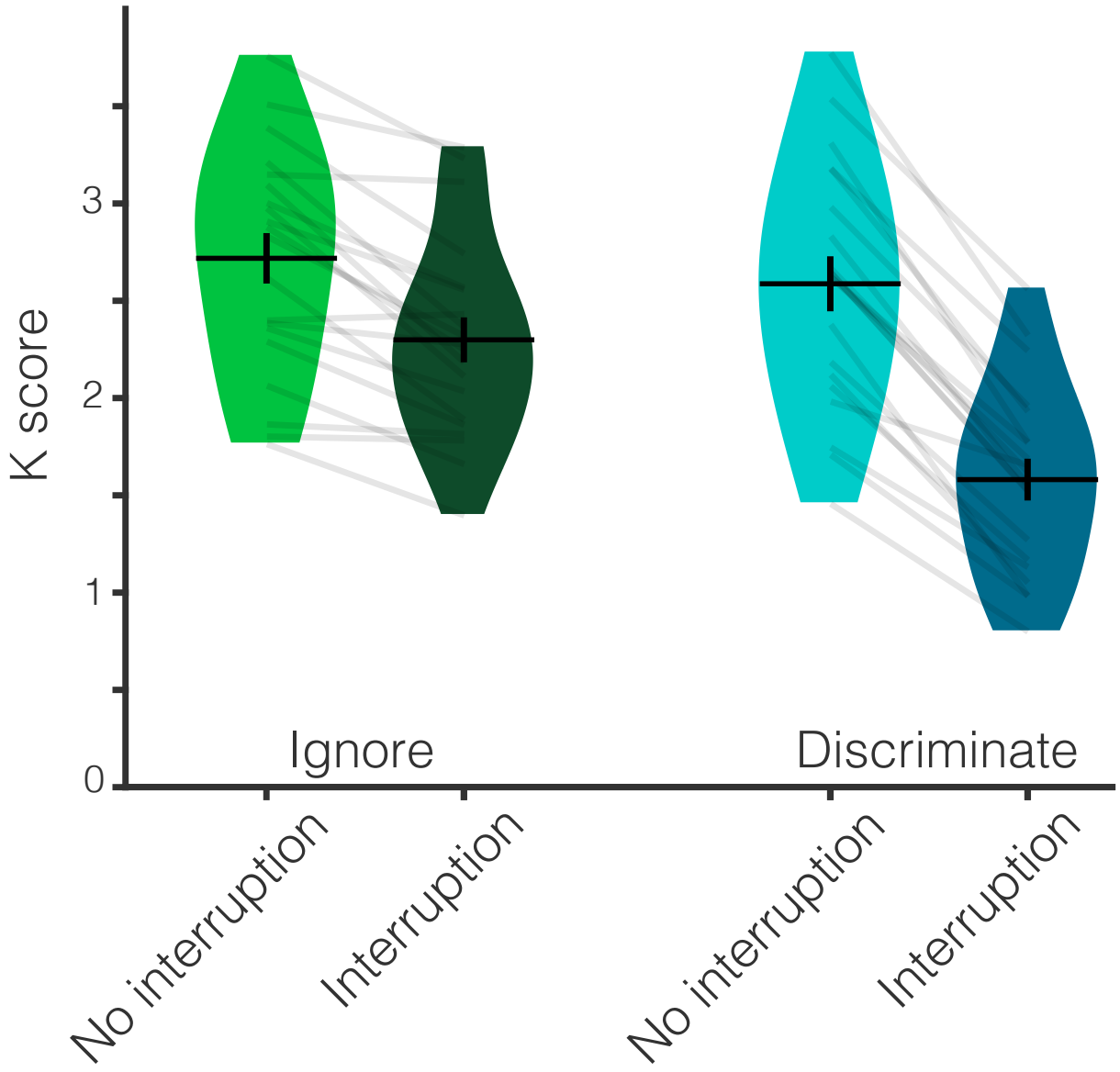

Figure 4-2. Behavioral results for Experiment 4-1.

(Figure 4-2 continued) Behavioral performance (K score) across the four conditions. Participants remembered fewer items when they were interrupted than when they were not interrupted. This

impact of interruptions was larger when participants had to discriminate the interrupter than when they could ignore them. Average K score is represented by the horizontal black line and the black error bars reflect the standard error of the mean. The distribution of K scores in each condition for all participants is represented by the violin plots. Light gray lines connect data from one participant across conditions.

Lateralized ERP

In this experiment, we analyzed lateralized alpha power, the contralateral delay activity (CDA), distractor positivity (P_D), and N2pc. For all of the below analyses, we calculated the difference between contralateral and ipsilateral activity (see Figure 4-3 for all neural results from Experiment 4-1, including average and standard error of the mean of the time course for each condition) and then took the difference (contralateral – ipsilateral). We then analyzed this difference value for the two conditions when interrupters were presented (relevant versus irrelevant) to determine whether there were any lateralized differences between conditions at each time window. To determine whether the signals were significantly lateralized, we additionally calculated one-way t-tests for each condition.

Pre-interrupter (400-650 ms). Before interrupter onset (400-650 ms), there were no lateralized ERPs in either condition (Ignore: $t(19)=-1.78$, $p=0.09$, $d=-0.40$; Discriminate: $t(19)=0.51$, $p=-0.62$, $d=0.11$), and there was no difference between trials with task-relevant versus task-irrelevant interrupters ($t(19)=-1.53$, $p=0.14$, $d=0.24$). This is what we expected because the memory array was presented centrally, and the lateralized interrupters had not yet appeared.

Post-interrupter I (850-950 ms). During this time window (equivalent to 200-300 ms post- interrupter, typically used for attention components), we observed a robust difference in lateralization between trials with task-relevant and task-irrelevant interrupters ($t(19)=3.19$, $p=0.005$, $d=0.71$). On trials with task-irrelevant interrupters, the lateralized ERP was positive

(M=0.24±0.45), and it was negative on trials with task-relevant interrupters (M=-0.44±0.98), suggesting the presence of an N2pc and P_D, respectively. To confirm this, we calculated the P_D and N2pc in specific a-priori time windows (see Methods for detail). One-way, one-sample t-tests against zero confirmed that there was a reliable P_D when task-irrelevant interrupters were presented (t(19)=3.19, p=0.002, d=0.71) and no P_D when task-irrelevant interrupters were presented (t(19)=-1.65, p=0.94, d=-0.37). Additionally, there was a reliable N2pc when task-relevant interrupters were presented (t(19)=-2.45, p=0.01, d=-0.55), but not when task-irrelevant interrupters were presented (t(19)=2.37, p=0.99, d=0.53)

Post-interrupter II (1050-1300 ms). In this time window, there was a significant difference in lateralization between trials with task-relevant and task-irrelevant interrupters (t(19)=4.32, p<0.001, d=0.97). The CDA was more lateralized when the task-relevant interrupters were presented (M=-0.59±0.55) than when task-irrelevant interrupters were presented (M=-0.08±0.30). In fact, the CDA was reliable on trials with task-relevant interrupters (one-sample: t(19)=-4.81, p<0.001, d=-1.08), but not on trials with task-irrelevant interrupters (one-sample: t(19)=-1.13, p=0.27, d=-0.25).

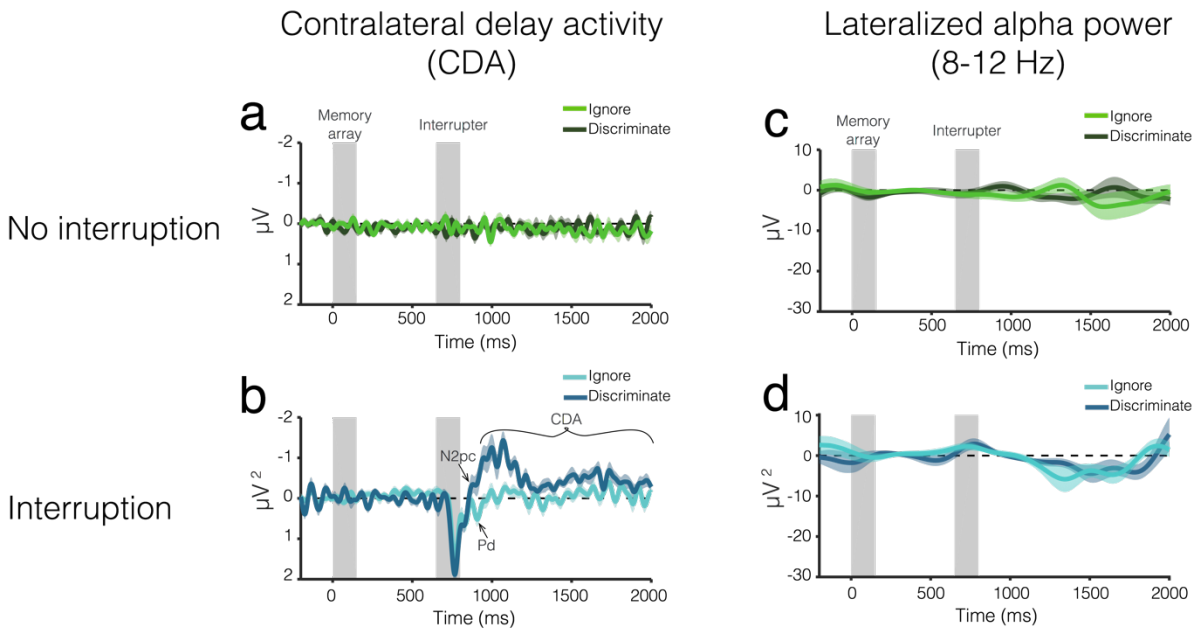

Figure 4-3. Contralateral delay activity and lateralized alpha power results for Experiment 4-1.

Average CDA amplitude over time for trials (b) with and (a) without interrupters. The light color envelopes around each line represent standard error of the mean for each condition. The first vertical gray bar (0-150 ms) represents when the memory array was on the screen, and the second gray bar (650-800 ms) represents when the interrupters were on the screen, if there were interrupters on that trial. The orange vertical bars represent the analyzed time windows (400-650 ms; 850-950 ms; and 1050-1300 ms). Lateralized alpha power over time for trials (d) with and (c) without interrupters.

Lateralized alpha power

Pre-interrupter (400-650 ms). Before interrupter onset (400-650 ms), alpha power was not significantly lateralized in either condition (Ignore: $t(19)=1.13$, $p=0.27$, $d=0.25$; Discriminate: $t(19)=-0.57$, $p=0.57$ $d=-0.13$), and there was no difference between trials with task-relevant versus task-irrelevant interrupters ($t(19)=0.991$, $p=0.33$, $d=0.22$). This was expected because the memory array was presented centrally, and the lateralized interrupters had not yet appeared.

Post-interrupter I (850-950 ms). Immediately following interrupter (850-950 ms), alpha power was not significantly lateralized in either condition (one-tailed t-tests against zero: Ignore: $t(19)=1.20$, $p=0.12$, $d=0.27$; Discriminate: $t(19)=1.50$, $p=0.08$, $d=0.33$), and lateralization did not vary between conditions ($t(19)=-0.23$, $p=0.82$, $d=-0.05$).

Post-interrupter II (1050-1650 ms). Towards the end of the trial (1050-1650 ms), alpha power was significantly lateralized in both conditions, consistent with a shift of spatial attention towards the interrupters (one-tailed tests against zero: Ignore: $t(19)=-2.33$, $p=0.015$, $d=-0.52$; Discriminate: $t(19)=-2.24$, $p=0.019$, $d=-0.50$). Interestingly, there was no difference in lateralization between the two conditions (paired-samples t-test: $t(19)=-0.89$, $p=0.40$, $d=-0.20$).

Conclusions

In Experiment 4-1, participants performed a WM change detection task with interrupters that appeared during the delay on a subset of trials. Behaviorally, participants remembered fewer items when they were interrupted than when they were not interrupted. This negative impact of interruption on behavior was larger when participants discriminated the task-relevant interrupters than when they ignored task-irrelevant interrupters.

The lateral position of the interrupters allowed us to assess how they were processed using a suite of lateralized ERP signals. Task-relevant interrupters elicited an N2pc followed by a sustained CDA when they had to be discriminated. This suggests that participants attend task-relevant interrupters, then encode them into visual working memory. Conversely, when task-irrelevant interrupters were presented, there was a P_D instead of an N2pc and no CDA. Thus, task-irrelevant interrupters were actively suppressed from being encoded into visual WM. In contrast, there is evidence that spatial attention may be captured regardless of whether the

interrupters were task relevant, as shown by a decline in alpha power contralateral to the position of the interrupters.

These findings suggest that observers could exert attentional control over whether the interrupters entered into working memory, and that this could be accomplished even when the interrupters captured spatial attention. Thus, these findings converge with prior work that has pointed towards distinct computational roles for CDA and alpha activity, with the former associated with item-based storage, and the latter associated with covert spatial attention (Gunseli et al., 2019; Hakim, Adam, et al., 2019).

Experiment 4-2

The relevance manipulation in Experiment 4-1 was a dual-task design, as it required participants to maintain information about two different tasks when task-relevant interrupters were presented. With this kind of manipulation, participants could always try to optimize their performance by trading off between the two tasks on some portion of trials, especially during the relevant condition blocks. It's possible that participants simply chose to utilize an "offline" strategy on some trials in which they did not attempt to actively maintain the target items so that they could dedicate resources to the discrimination task. This strategy may be less likely in the irrelevant condition when subjects knew they do not need to do anything with the interrupter items. Such a difference in strategy could plausibly explain why we observe a CDA to the task-relevant interrupters and not for the task-irrelevant interrupters. It is also generally consistent with our finding that the behavioral deficit was largest for the relevant condition. If this were the case, it would suggest that the results of Experiment 4-1 were the result of general strategic differences between the conditions that occur prior to interrupter rather than the impact of the

relevant interrupters themselves. Therefore, in Experiment 4-2, we tested whether participants in the relevant condition actively encoded the target items into working memory prior to the onset of the interrupters, or whether they chose not to actively encode or maintain the target array in anticipation of making the discrimination. Therefore, the key question in Experiment 4-2 is whether there are differences in the CDA and alpha power lateralization between the ignore and discriminate conditions during the pre- interrupter period. If participants did not actively store the memory array items in WM pre- interrupter, the CDA should be reduced or eliminated in the relevant condition as compared to the irrelevant condition.

Materials & Methods

Participants. Twenty-nine volunteers, naïve to the objective of the experiment participated for payment ($15 USD per hour). The data from 8 participants were excluded from the analysis because of too many artifacts (same criteria as Experiment 4-1). The remaining 21 participants (9 male) were between the ages of 30 and 18 (M = 22.1, SD = 3.9). We recruited participants from the same subject pool as Experiment 4-1, and we permitted participants from Experiment 4-1 to participate in Experiment 4-2. In total, 4 participants completed both experiments.

Apparatus and stimuli. The apparatus was identical to Experiment 4-1. Stimuli (Figure 4-4) were also identical to Experiment 4-1 with the following exceptions. We were interested in the neural representations of the memory array items. Therefore, we presented the memory array items laterally and the interrupters centrally. Thus, CDA amplitude can be interpreted as encoding and maintenance of the memory array and lateralization of alpha power can be interpreted as a shift of attention towards laterally presented memory items. Therefore, at the beginning of the experiment, a horizontal diamond comprised of a green (RGB = 74, 183, 72;

52.8 cd/m^2) and a pink (RGB = 183, 73, 177; 31.7 cd/m^2) triangle appeared on the vertical midline 0.65° above the fixation dot. In 50% of the trials, the pink triangle pointed to the left side and the green triangle pointed to the right side, in the remaining 50% of the trials this was inverse. Half the participants were instructed to attend the hemifield that the pink triangle pointed to, and the other half was instructed to attend the hemifield to which the green triangle pointed. Memory displayed showed an array of 3 colored squares in each hemifield. Within each hemifield, there were one or two squares in the upper quadrant and two or one square in the lower quadrant. Squares could appear within an area of the display subtending 6° to the left or right of fixation and 3.1° above and below fixation. The interrupters display showed four colored squares of the same size as the ones from the memory display along the midline of the screen, drawn from the remaining colors. These interrupting items were shown on the vertical midline with a randomly jittered horizontal offset of maximally 0.55° (half of an object). Probe displays showed one colored square in each hemifield in the same location as one of the squares, randomly picked, in the original array. The color of the square in the unattended hemifield was the same as the original square on 50% of trials, and different on the other 50% of trials.

Procedure. The procedure was identical to Experiment 4-1 with the following exception. Each trial began with a cue display (500 ms) indicating the to-be-attended side of the screen (left or right).

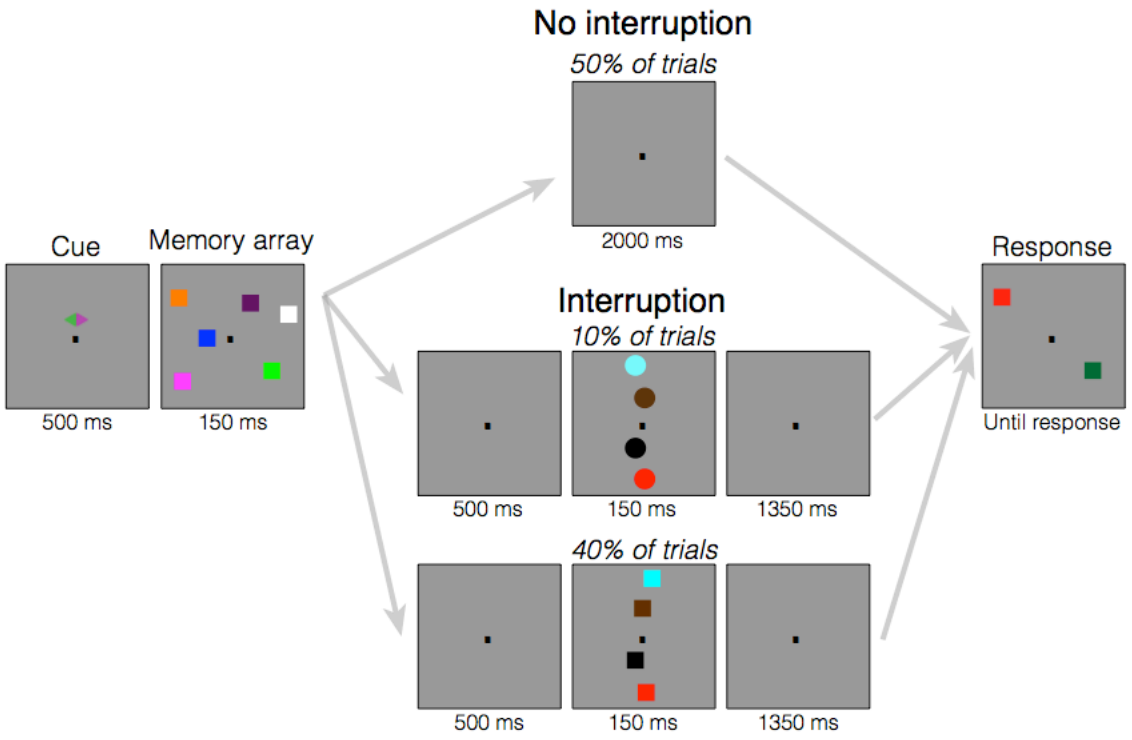

Figure 4-4. Task design for Experiment 4-2.

At the start of each trial, a cue appeared above the fixation dot, which indicated to participants which side of the screen they should attend. Participants either attended to the green or purple side (counterbalanced across participants). Following the cue, the memory array appeared, which consisted of three colored squares on each side of the screen. Participants were told to remember the colors of the squares on the cued side. Following the memory array, the screen went blank. Then, either the screen remained blank the entire delay (no interrupter conditions) or the screen went blank for 500 ms (interrupter conditions) followed by a series of four objects (circles or squares) along the midline. When participants were in the "Ignore" condition, they were told to always ignore these interrupter objects. When they were in the "Discriminate" condition, they were told to determine the shape of the stimuli in order to report whether the stimuli were circles. They were told to withhold their response until the response screen appeared. Following interrupters, the screen then went blank for the rest of the delay. On the final screen, one square on either side of the screen re-appeared and could either be the same color as the original square or it could be a different color that did not appear in the display. In both conditions, participants had to report whether the square on the cued side of the screen changed colors. In the "Discriminate" condition, participants additionally had to report whether the interrupter objects were circles, if there were interrupter on that trial.

Artifact rejection & analyses. Artifact rejection and analyses were identical to Experiment 4-1 with the following exceptions. Since circle interrupters were only 10% of

interrupter trials and required an additional response, we only included square interrupter trials in all analyses. Additionally, we analyzed EEG data in all four conditions because stimuli were presented laterally in all cases. We compared conditions with a repeated measures ANOVA with the within-subjects factors Interrupter (interrupters versus no interrupters) and Relevance (ignore versus discriminate). For alpha power, we analyzed all of the same time windows as the CDA because previous research has directly investigated the time course of alpha lateralization following centrally presented interrupters (Hakim et al 2020). Finally, we did not analyze the P_D and N2pc because interrupters were presented centrally in this experiment. We corrected all of the ANOVA results in the two post- interrupter time windows for multiple comparisons using a Bonferroni correction for two comparisons (significance threshold: $p<0.025$).

Results

Behavior

Behavioral performance was significantly above chance in all conditions (all one-sample t-tests, $p\leq0.001$). Participants remembered fewer items when they were interrupted ($M=1.62\pm0.66$) than when they were not interrupted ($M=1.94\pm0.65$), significant main effect of Interruption ($F(1,20)=100.21$, $p<0.001$, $\eta_p^2=0.83$). Participants also remembered fewer items when they had to discriminate the interrupters ($M=1.69\pm0.64$) than when they could ignore the interrupters ($M=1.87\pm0.63$), significant main effect of Relevance ($F(1,20)=10.20$, $p=0.005$, $\eta_p^2=0.34$). The difference between trials with and without interrupters was significantly larger when the task-relevant interrupters had to be discriminated ($\Delta M=0.479\pm0.221$) than when the task-irrelevant interrupters could be ignored ($\Delta M=0.177\pm0.258$), significant interaction of Interruption and Relevance ($F(1,20)=13.67$, $p=0.001$, $\eta_p^2=0.406$). The significant follow-up t-test

showed that the difference between trials with and without interrupters was significantly larger in the Ignore than the Interruption condition (t(20)=-3.698, p=0.001). Additionally, participants were 93.6% (±3.5%) accurate on the interrupter task. Behavioral results from Experiment 4-2 are depicted in Figure 4-5.

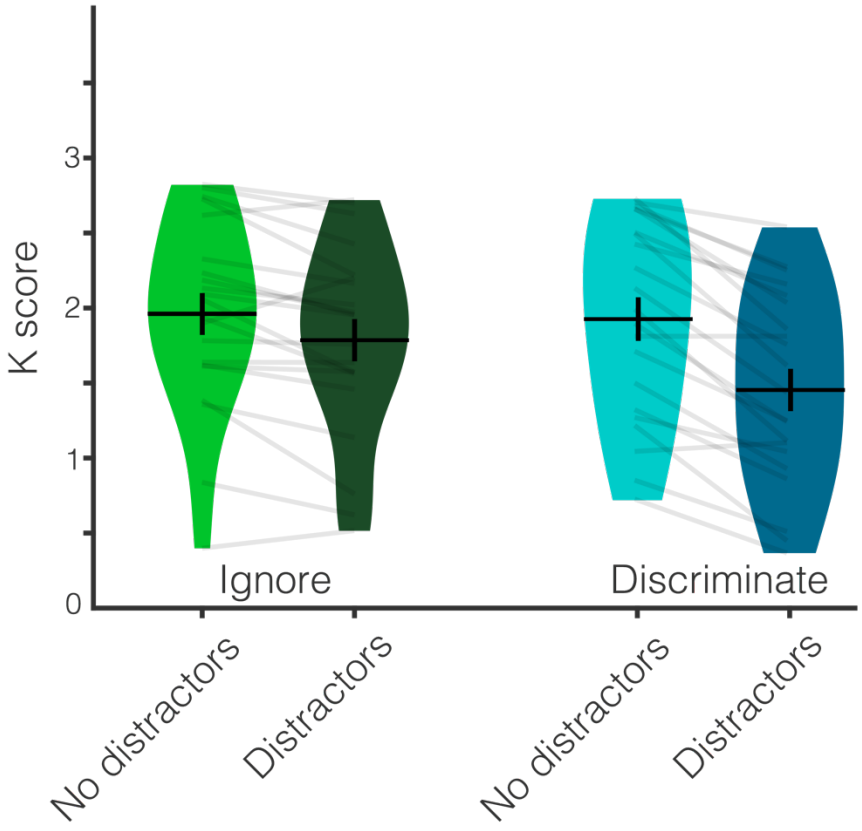

Figure 4-5. Behavioral results for Experiment 4-2.

Behavioral performance (K score) across the four conditions. Participants remembered fewer items when they were interrupted than when they were not interrupted. This impact of interruption was larger when participants had to discriminate the interrupters than when they could ignore them. Average K score is represented by the horizontal black line and the black error bars reflect the standard error of the mean. The distribution of K scores in each condition for all participants is represented by the violin plots. Light gray lines connect data from one participant across conditions.

Lateralized ERP

Just as in Experiment 4-1, we ran a repeated measures ANOVA with the factors Interruption (no interrupter, interrupter) and Relevance (ignore, discriminate) to determine whether there were any differences between conditions at each time point. In this experiment, we analyzed lateralized alpha power and CDA. To determine whether the signals were significantly lateralized, we additionally calculated one-way t-tests for each condition. Results displayed in Figure 4-6 (figure includes the average and standard error of the mean of the time course for each condition).

Pre-interruption (400-650 ms). The CDA was reliable in all conditions (all one-sample t-tests $p \leq 0.001$), and CDA amplitude did not vary between conditions (main effect of Interruption and Relevance and their interaction all $p \geq 0.24$).

Post-interrupter I (850-950 ms). CDA amplitude was larger in the ignore condition (M=-0.90±0.76) than in the discriminate (M=-0.59±0.79), regardless of whether interrupters appeared (significant main effect of Relevance: $F(1,20)=2.08$, $p=0.02$, $\eta_p^2=0.23$). The main effect of Interruption and the interaction of Interruption and Relevance were not significant (both $p \geq 0.12$).

Post-interrupter II (1050-1300 ms). CDA amplitude was larger on trials without interrupters (M=-0.83±0.51) than on trials with interrupters (M = -0.22 ± 0.73) regardless of relevance (significant main effect of Interruption: $F(1,20) = 17.73$, $p < 0.001$, $\eta_p^2 = 0.47$). The main effect of Relevance was trending, but not significant ($F(1,20) = 3.65$, $p=0.07$, $\eta_p^2=0.15$), and the interaction of Interruption and Relevance was not significant ($F(1,20)=0.29$, $p = 0.69$, $\eta_p^2 =$

0.01). When interrupters were not present, there was a reliable CDA regardless of condition (both $p \leq 0.001$).

Lateralized alpha power

Pre-interruption (400-650 ms). Before the interrupters appeared (400-650 ms), alpha power was significantly lateralized in all conditions (all $p \leq 0.004$), and lateralization did not vary between condition (main effect of Interruption and Relevance and their interaction were not significant all $p \geq 0.25$).

Post-interrupter I (850-950 ms). Immediately following interrupters (850-950 ms), alpha power was more lateralized on trials without interrupters (M=-25.39±33.33) than on trials with interrupters (M=-3.71±6.89) regardless of relevance (significant main effect of Interruption: $F(1,20)=9.37$, $p=0.006$, $\eta_p^2=0.32$). The main effect of Relevance and the interaction of Relevance and Interruption were not significant (both $p \geq 0.42$).

Post-interrupter II (1050-1300 ms). Toward the end of the delay (1050-1300 ms), alpha power was more lateralized on trials without interrupters (M=-13.66±24.62) than on trials with interrupters (M=-3.77±12.11; significant main effect of Interruption: $F(1,20)=5.98$, $p=0.02$, $\eta_p^2=0.23$), regardless of relevance. The main effect of Interruption and the interaction of Relevance and Interruption were not significant (both $p \geq 0.98$). Follow-up one-sample t-tests revealed that alpha power was significantly lateralized on trials without interrupters (both $p \leq 0.03$), but not on trials with interrupters (both $p \geq 0.09$), regardless of relevance. To summarize, the presentation of the central interrupters disrupted alpha lateralization towards the target locations, and this effect did not depend on whether the interrupters were task relevant.

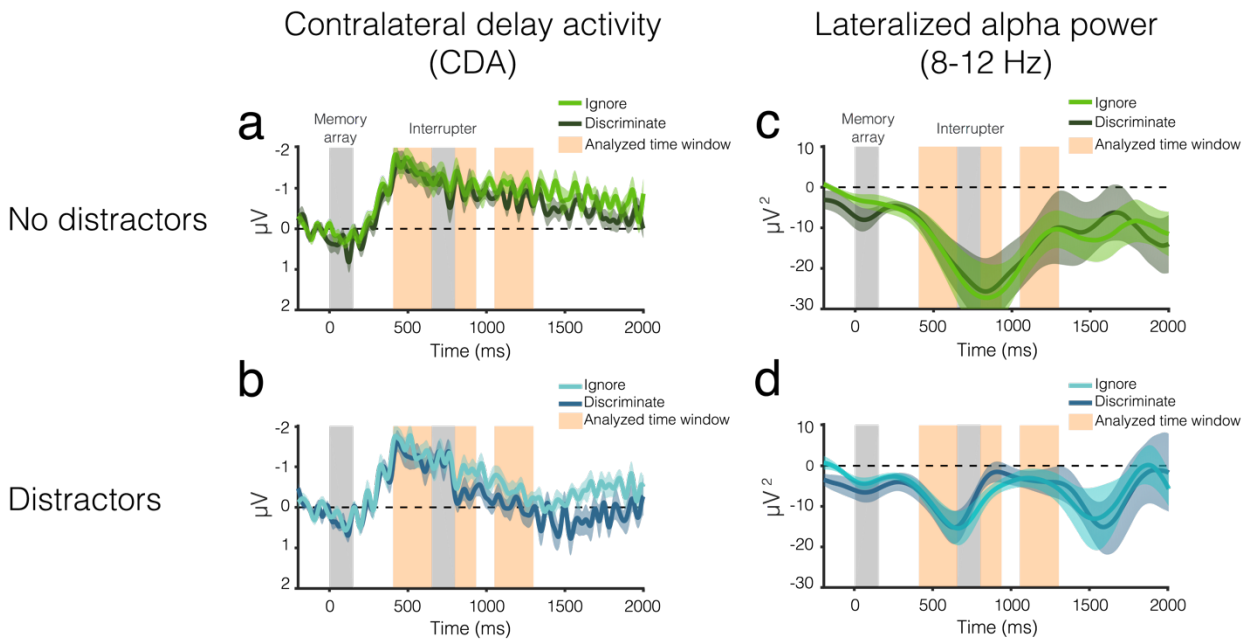

Figure 4-6. Contralateral delay activity and lateralized alpha power results from Experiment 4-2.

CDA amplitude over time for trials (b) with and (a) without interrupters. The light color envelopes around each line represent standard error of the mean for each condition. The first vertical gray bar (0-150 ms) represents when the memory array was on the screen, and the second gray bar (650-800 ms) represents when the interrupters were on the screen, if there were interrupters on that trial. The orange vertical bars represent the analyzed time windows (400-650 ms; 850-950 ms; and 1050-1300 ms). Lateralized alpha power over time for trials (d) with and (c) without interrupters.

Conclusions

In Experiment 4-2, we were interested in whether the relevance of the interrupters affected the likelihood of encoding and maintaining target information in WM. The relevant interrupter condition in our two experiments is similar to a dual-task design as participants have to maintain information about two separate tasks. Participants could decide to drop information about the initial target array in order to encode the new interrupting information from the second task. Alternatively, they could attempt to sustain the initial working memory representations at

the expense of sufficiently attending the relevant interrupters. How did participants address this trade-off between tasks in the present study?

Before the interrupters appeared, there was no difference in CDA amplitude or alpha power lateralization between the ignore and discriminate conditions. There was a clear CDA and alpha power lateralization for both conditions, with no differences between them. This indicates that in both conditions, participants encoded and maintained lateralized working memory representations and sustained their attention. Differences between conditions only emerged after the time when interrupters were supposed to appear. Therefore, the results from Experiment 4-1 are not simply due to a dual-task tradeoff between the encoding and maintenance of the memory array and the interrupters.

The interrupters always appeared at the same time during the delay. Therefore, participants could anticipate when they might be interrupted. Around the time when interrupters typically appeared, the CDA was smaller when participants anticipated a relevant interrupter than when they anticipated a task-irrelevant interrupter. During this same time window, alpha power lateralization depended on the presence of interrupters. Towards the end of the trial, however, both CDA amplitude and alpha power lateralization were closer to baseline when interrupters were present than when they were not, regardless of relevance. Overall, when participants anticipate that they may have to integrate new information into working memory, they hold less information about the target array around the time that they think new information will appear, even if this new information doesn't appear. However, toward the end of the trial, relevance no longer impacts the likelihood of sustaining attention or maintaining information about the target array. Overall, we can rule out the alternate explanation of experiment 4-1 that participants prematurely drop information about the target array when relevant interrupters are

presented. Participants were clearly attending and maintaining memoranda even when relevant interrupters appeared.

Discussion

The key finding of the present study was that processing of stimuli that disrupt ongoing WM representations depended on their relevance. Task-relevant interrupters were encoded and maintained in WM, whereas task-irrelevant interrupters were suppressed and never entered WM. On the other hand, spatial attention was captured regardless of stimulus relevance. In Experiment 4-2, we investigated whether participants in the relevant condition of Experiment 4-1 actively encoded and maintained memory items prior to the onset of the interrupters. We found that pre-interrupter, there were no differences between conditions. Participants encoded and maintained working memory representations and sustained attention to the memory items equally in both conditions. Differences between conditions only emerged after interrupter onset, indicating that the results from Experiment 4-1 are not solely driven by a dual-task tradeoff between the maintenance of the memory array and the interrupters.

We observed distinct effects of interruption on lateralized ERP signals and lateralized alpha power in these experiments. The results converge with past proposals of a distinction between item-based and spatial capture of attention (Hakim, Adam, et al., 2019; Hakim et al., 2020). In our procedure, alpha oscillations showed that interrupters captured spatial attention, regardless of task relevance. By contrast, the formation of item-based representations in working memory was completely determined by task relevance, such that encoding into working memory was suppressed when observers could ignore the interrupters. Interestingly, both item-based

storage and spatial attention towards the memoranda were eventually disrupted by the presentation of the interrupters, regardless of whether a dual task was imposed.

Spatial capture by task-relevant and task-irrelevant interrupters

In Experiment 4-1, we directly investigated how interrupters are processed. In this experiment, we found that alpha power was significantly lateralized following the onset of both task-relevant and task-irrelevant interrupters. This result provides evidence that spatial attention shifted to the location of both the relevant and irrelevant interrupters. We designed our interrupters to be very similar to the original memoranda in order to induce a large behavioral deficit on interrupter trials compared to non- interrupter trials. Previous research has shown that when interrupters share features with memoranda (Hollingworth & Beck, 2016; Olivers & Eimer, 2011; Soto et al., 2008; van Moorselaar et al., 2015) or are part of the attentional set (Folk et al., 2008; Folk & Remington, 1998), spatial attention can be captured at the location of the interrupters. However, the similarity between targets and interrupters is a continuum (Duncan & Humphreys, 1989) and attentional capture increases with target- interrupter similarity (Ansorge & Heumann, 2003; Ludwig & Gilchrist, 2002). In the present study, relevant interrupters were the same shape as the targets and drew their color from the same pool of colors as the targets, i.e., they had potential target colors. However, no interrupters were ever the same color as a target within a trial. Nevertheless, the finding that spatial attention was captured by both types of interrupters may be partially due to the fact that even irrelevant interrupters were perceptually similar enough to the targets (i.e., contingent capture). Future research should determine whether spatial attention is necessarily captured when interrupters do not share any similarities with currently maintained working memory representations. Understanding whether spatial and item-

based attention are captured along the entire continuum is an important question as it will give insight into the question of how features are weighted in WM and how this affects attention deployment.

Voluntary control of item-based capture

In Experiment 4-1, irrelevant interrupters elicited a Pd, suggesting that they were actively suppressed. Relevant interrupters, however, first elicited an N2pc which indicates that these items were individuated. The subsequent lateralized negativity then transitioned into a CDA, suggesting that interrupters were not only individuated but also encoded into WM. These results illustrate that participants dynamically respond to task demands by suppressing irrelevant interrupters from WM and only encoding relevant interrupters into working memory. Even when salient stimuli captured spatial attention, participants still had voluntary control over whether to store that information in working memory. These results are in line with previous findings showing that successful suppression of irrelevant information can contribute to better performance (Feldmann-Wüstefeld et al., 2016; Gaspar & McDonald, 2014a; Sawaki & Luck, 2012; Weaver et al., 2017). For example, when target identity is correctly reported in a visual search task, a concurrently presented salient interrupter elicits a pronounced Pd component, indicative of active suppression (Feldmann-Wüstefeld et al., 2020). Conversely, when the interrupter identity is erroneously reported, interrupters elicit a CDA and a less pronounced Pd, suggesting that the interrupter was encoded into WM.

Our results also nicely align with the working memory gating literature. This literature provides a framework to explain which information is allowed to enter working memory and which information is blocked (Badre, 2012; Chatham et al., 2014; Chatham & Badre, 2015;

O'Reilly & Frank, 2006). According to this account, the working memory gate is the mechanism by which irrelevant information is blocked from entering. When the working memory gate is open, it allows information to enter working memory. When it is closed, ongoing working memory representations are sustained, while irrelevant information is blocked (Badre, 2012). The working memory gating literature has mostly used fMRI to demonstrate which parts of the brain are involved in working memory gating and maintenance and has not distinguished between WM gating and capture of spatial attention. Our results suggest that we may be able use EEG activity to track working memory gating as well. We propose that the CDA tracks how much information passes through the gate, whereas the Pd reflects the gate itself. Previous research has shown that Pd amplitude scales with the number of items that were blocked from entering working memory (Feldmann-Wüstefeld et al., 2019). Therefore, Pd amplitude may reflect how firmly the working memory gate was closed. Conversely, the CDA could reflect how much information is encoded into working memory, with relevant information more likely to pass through the gate. Our results suggest that how firmly the gate is closed can be controlled in a top-down manner whereas other factors, such as physical salience, determine how much information enter working memory. Future research could investigate the precise temporal dynamics of working memory input and output gating using these proposed EEG signals. For example, if the working memory gate accidentally allows irrelevant information into working memory, how long is that information maintained in working memory before it is dropped?

Is attentional capture obligatory?

Our findings provide a new framework in which we can investigate attentional capture. We propose that attentional capture is comprised of item-based and spatial capture. Item-based

capture involves forming working memory representations of the new stimuli. Based on our findings, item-based capture appears to be subject to voluntary attentional control. It allows relevant stimuli to enter working memory, while irrelevant stimuli are suppressed. We also have clear evidence that, in our specific task context, spatial capture occurred when interrupters were present, regardless of top-down goals. However, whether spatial capture merely happened because of perceptual salience (i.e. bottom-up capture) or whether similarity with task-relevant items (continent capture) contributed to attention deployment is unclear and should be the focus of future research. Thus, WM gating can successfully block irrelevant information, even when spatial attention is captured. From our perspective, information needs to be held in working memory in order for it to be processed. Therefore, if interrupting stimuli are not encoded into working memory via item-based capture, they are not fully processed, even if they captured spatial attention. Selecting an item using spatial attention is necessary, but not sufficient, to encode it into working memory.

We investigated item-based and spatial capture during maintenance of ongoing working memory representations. However, the majority of the attentional capture literature has investigated this process during encoding. Therefore, future research could apply this new framework of attentional capture to investigate item-based and spatial capture during encoding. This could potentially provide insight into the ongoing debate about whether attentional capture during encoding is obligatory (Feldmann-Wüstefeld et al., 2020; Feldmann-Wüstefeld & Schubö, 2013; Gaspar & McDonald, 2014b; Gaspelin et al., 2015, 2017; Gaspelin & Luck, 2018; Hickey et al., 2009; Liesefeld et al., 2017; Sawaki et al., 2012). We hypothesize that the ways in which the two sub-components of attentional capture (item-based and spatial capture) respond during encoding should be similar to how they respond during maintenance. That is, we

hypothesize that during encoding, spatial capture may happen regardless their relevance, whereas item-based capture may be subject to voluntary attentional control.

To further elucidate the neural mechanism underlying working memory performance, future research could also compare the EEG signal for accurate versus inaccurate responses. In the current study, behavioral performance was very high (mean accuracy of 79.9% and 78.7% in Experiment 4-1 and 4-2), which means that an insufficient number of trials was available for a reliable comparison. A direct comparison of correct and incorrect trials could reveal which cognitive mechanism are involved in WM failures. For example, previous research found that the P_D component is smaller on incorrect than on correct trials in a visual search task (Feldmann-Wüstefeld et al., 2020). For the present research question, it could be particularly insightful to compare CDA and alpha-band activity between correct and incorrect trials to reveal whether item-based or spatial attention contribute more to behavioral errors.

Impact of interrupters on ongoing working memory representations

Previous research has demonstrated that salient interrupter stimuli interfere with object representations in WM and cause attention to shift away from maintained representations (Hakim, et al., 2019). Our work replicates and extends these findings by adding a top-down perspective. Participants had to either attend (relevant) or ignore (irrelevant) interrupters. Here, we show that the CDA, a neural measure of working memory load, was initially influenced by top-down goals. When participants anticipated that they may have to encode additional information into WM, the CDA was smaller than when participants knew that interrupter stimuli could be ignored. However, the CDA was at baseline toward the end of trials that contained both relevant and irrelevant interrupters. This suggests that both types of interrupters harmed

lateralized object representations of the memoranda. However, towards the end of the trial, alpha power, a neural measure of spatial attention, shifted to baseline following both relevant and irrelevant interrupters, suggesting that participants shifted their attention away from the locations of the original memoranda regardless of whether the interrupter information was relevant. These results illustrate that the observers' goals determine the encoding of item-based representations, even when spatial attention is captured. Toward the end of the trial, however, interrupters harm both spatial attention and the object representations of the memoranda, regardless of top-down goals.

Sub-component processes of attentional capture

Our two experiments provide evidence that attentional capture may be comprised of at least two distinct sub-component processes: item-based capture and spatial capture. Future research should investigate how these sub-processes interact with other forms of attention, such as sustained attention and vigilance. For example, previous research has found that participants with lower executive WM capacity tend to mind-wander more than those with higher executive WM capacity (Kane et al., 2007; McVay & Kane, 2009, 2010). Does an analogous relationship exist between spatial attention and mind-wandering? Additionally, if a participant is in a high attentional state, is their spatial attention less likely to be captured? Thus far, research on sustained attention, spatial attention, and mind-wandering have largely developed in parallel. The type of experimental design that we used in the current study may provide a fruitful avenue to integrate these currently disparate views of attention.

Conclusions

Previous research has treated attentional capture as a monolithic process. Here, we present new evidence that there are at least two sub-component processes of attentional capture that are neurally dissociable: spatial capture and item-based capture. Lateralized alpha power indexes spatial capture, a process that involves a shift of spatial attention. By contrast, item-based capture is tracked by the N2pc and CDA when item-based representations are deemed relevant and allowed to enter working memory, while the P_D tracks the active suppression of items from WM. This fractionation of attentional capture into distinct sub-component processes provides a framework by which the fate of ongoing WM processes after interrupters can be explained. We show that relevant interrupters trigger both of these dissociable processes. Irrelevant interrupters, however, only trigger spatial capture.

PART III

Predicting individual differences in working memory capacity

CHAPTER 5.

Neural activity can be used to predict working memory capacity in unseen individuals

Human brains share a common template of functional organization. Nearly every person, for example, shows a retinotopic map in primary visual cortex (Engel et al., 1997) and a face-sensitive region in inferior temporal cortex (Kanwisher et al., 1997). Electroencephalogram (EEG) activity signatures of the number of items in working memory are reliable enough to be detected at the single-subject level (Vogel & Machizawa, 2004b). Even in the absence of an explicit task, individuals show synchronous activity in a stereotyped set of brain networks, such as the default mode (Damoiseaux et al., 2006) and frontoparietal (Duncan & Owen, 2000) networks.

Atop this shared organizational template is significant individual idiosyncrasy in functional brain architecture (Charest et al., 2014). Functional MRI studies have revealed that each person has a unique pattern of functional connectivity, or correlated activity between spatially distinct brain regions, that distinguishes them from others and remains stable across cognitive states (Finn et al., 2015; Gratton et al., 2018; Miranda-Dominguez et al., 2014). Furthermore, these unique patterns, or whole-brain connectomes, appear cognitively meaningful, predicting individual differences in behaviors including fluid intelligence (Finn et al., 2015; Greene et al., 2018) and attention (Kessler et al., 2016; O'Halloran et al., 2018; Poole et al., 2016; Rosenberg et al., 2016).

Brain-based predictive models rely on this interesting duality in brain function: broadly similar organization with consequential individual variation. That is, a neural system common across individuals is necessary to build brain-based biomarkers because if every individual relied

on a different neural system to achieve a particular behavior, predictive models would fail to generalize across people. However, systematic idiosyncrasies in these common brain networks are what allow models to predict each person's unique set of cognitive abilities. Without differences in these common neural systems, predictive models would fail to differentiate individuals. This duality in brain function has primarily been observed when using techniques, such as MRI, that consider the spatial organization of the brain. In the current work, we seek to address the open theoretical question of whether these common and unique signals of cognition are also present in temporally sensitive, but spatially insensitive, neural signals, such as EEG.

Based in part on evidence from connectome-based models of behavior, work in cognitive and network neuroscience has argued that cognition relies on coordinated activity in large-scale, high-density brain networks (Bressler & Menon, 2010; Medaglia et al., 2015; Park & Friston, 2013). This may suggest that dense functional networks revealed by techniques with relatively high spatial resolution, such as fMRI, are uniquely informative of behavior. This hypothesis has not been directly tested. However, studies relating cognitive abilities to patterns of statistical dependence between electrical or magnetic signals (i.e., measured using EEG or magnetoencephalography [MEG]; Burgess & Ali, 2002; Damaševičius et al., 2018; Fellrath et al., 2016; Karamzadeh et al., 2013; Nentwich et al., 2020; Palva et al., 2010)—which often include fewer than one percent of the spatial resolution typically available to fMRI functional connectivity models—suggest that cognitively meaningful variability in brain function may be sampled using sparse, high-frequency neural signals. In this paper, we directly test the hypothesis that dense functional networks are uniquely informative of behavior. To do so, we ask whether inter-electrode correlations measured using EEG can 1) identify individuals and 2) predict trait-like cognitive abilities across individuals from completely independent datasets.

In addition to open theoretical questions about brain function and cognition, obstacles to widespread adoption of existing brain-based models of behavior remain. First, the majority of MRI-based models have not tested the generalizability of their predictions to unseen individual and datasets (Woo et al., 2017). This limits our ability to draw conclusions about their robustness and replicability (Poldrack et al., 2020). Second, work has suggested that, in some cases, confounds including head motion can influence observed relationships between fMRI networks and behavior (Siegel et al., 2017). Finally, the costs of MRI for researchers, clinicians, and participants has so far limited translation to real-world settings.

Here we address these open theoretical questions and practical challenges by using a direct measure of neural activity that is easy and affordable to implement: EEG. Across two EEG datasets, each with 165+ individuals, collected at different universities with different EEG systems (passive vs. active), we show that inter-electrode correlations between trial-evoked ERPs are unique across individuals and stable within individuals. We next demonstrate that models based on sparse inter-electrode correlations generalize across individuals and independent datasets to predict individual differences in working memory capacity, a critical cognitive ability. Finally, we show that the same set of inter-electrode correlations ("EEG network") that predicts working memory capacity predicts general fluid intelligence in novel individuals. Thus, sparse EEG networks reveal a signature of trait-like cognitive abilities in humans and provide a new, more affordable and accessible approach for predicting cognitive ability from brain function.

Materials & Methods

Data were collected at two different study sites from individuals of both sexes: the University of Oregon Eugene and the University of Chicago. Whereas data from the University of Oregon were collected as part of a single study, data from the University of Chicago were compiled from multiple studies. Participants were recruited from the respective university network and from their surrounding communities. Some of these samples are described in previous publications, while others are unpublished. All data are described below.

Participant information.

University of Oregon. Data collected at Oregon were part of one study, and the results from this dataset were previously published (Unsworth et al., 2015). All participants (107 female) gave written informed consent according to procedures approved by the University of Oregon institutional review board. Participants were compensated for participation with course credit or monetary payment ($8/hr for behavior, $10/hr for EEG). Our analyses include data from 171 individuals. The numbers of participants included here is different than the number of participants included in the original study because we only required that participants have usable EEG and change detection data, whereas the prior analyses required participants to have completed all tasks.

University of Chicago. Data from the University of Chicago were collected by multiple experimenters for multiple independent studies. Experiments were selected for inclusion based on whether they included a lateralized change detection task. All lateralized change detection

EEG experiments that have been run by the Awh/Vogel lab at the University of Chicago task were included.

Experimental procedures were approved by The University of Chicago Institutional Review Board. All participants gave informed consent and were compensated for their participation with cash payment ($15 per hour); participants reported normal color vision and normal or corrected-to-normal visual acuity. For the current analyses, University of Chicago data were combined into one large sample. Of note, some individuals participated in more than one University of Chicago experiment. For these individuals, behavioral data and EEG signal time-series for all trials from all of the experiments in which they participated were averaged. Across all of the Chicago studies, there were 165 unique individuals. For a subset of the Chicago experiments, gender information is currently inaccessible due to Covid-19 limitations. Therefore, only the published studies contain information about gender.

Chicago study #1 was previously published (Hakim, Adam, et al., 2019) and includes data from four separate experiments: 28 participants (13 female) in experiment 1a, 20 participants (10 female) in experiment 1b, 20 participants (10 female) in experiment 2a, and 29 participants (13 female) in experiment 2b. Chicago study #2 was previously published (Hakim, Feldmann-Wüstefeld, et al., 2019) and includes data from two separate experiments: 20 participants (8 female) in experiment 1 and 20 participants (9 female) in experiment 2. Chicago study #3 was previously published (Hakim et al, 2020b) and contains data from two separate experiments, however, we only analyzed data from experiment 1, which contained data from 21 participants (15 female) because this was the only experiment that included a lateralized change detection task. Chicago study #4 is not published and contains data from one experiment (20 participants). Chicago study #5 is not published and contains data from two experiments (20

participants in experiment 1 and 19 participants in experiment 2). Chicago study #6 is not yet published and contains data from two experiments (25 participants in experiment 1 and 19 participants in experiment 2).

Experimental Design. In all experiments, participants performed a lateralized change detection task. At the beginning of each trial, a central cue was presented to indicate which side (left or right) of the screen to pay attention to (Chicago study #6 did not have a cue on a subset of trials. See below for further information.) Following the arrow cue, a memory array appeared, which consisted of a series of objects on both sides of the screen. Participants were instructed to remember the objects that were presented on the cued side of the screen, while ignoring the objects on the other side. Following the memory array, there was a blank retention interval. The exact duration of the retention interval varied across experiments. The response screen then appeared, which consisted of one object on each side of the screen. Participants had to indicate if the object on the cued side was identical to the original object that was presented at that location (Chicago study #5 had a two-alternative forced choice response. See below for further information.) The exact duration and stimulus parameters varied across experiments. For published datasets, these details can be found in the original publications (Hakim, Feldmann-Wüstefeld, et al., 2019, 2019; Unsworth et al., 2015). For unpublished datasets, details are described below.

University of Chicago study #4. Stimuli were presented on a 24-in. LCD computer screen (BenQ XL2430T; 120-Hz refresh rate) on a Dell Optiplex 9020 computer. Participants were seated with their heads on a chin rest 74 cm from the screen. Each trial began with a blank

inter-trial interval (750 ms), followed by a diamond cue (500 ms) indicating the relevant side of the screen (right or left). This diamond cue (maximum width = 0.65°, maximum height = 0.65°) was centered 0.65° above the fixation dot and was half green (RGB value: = 74, 183, 72) and half pink (RGB value: = 183, 73, 177). Half of the participants were instructed to attend to the green side, and the other half were instructed to attend to the pink side. After the cue, colored squares (1.1° × 1.1°) briefly appeared in each hemifield (150 ms) with a minimum of 2.10° (1.5 objects) between each square. Four colored squares appeared on each side of the screen, and then disappeared for 2000 ms. Squares could appear within a subset of the display subtending 3.1° to the left or right of fixation and 3.5° above and below fixation. Colors for the squares were selected randomly from a set of nine possible colors (RGB values: red = 255, 0, 0; green = 0, 255, 0; blue = 0, 0, 255; yellow = 255, 255, 0; magenta = 255, 0, 255; cyan = 0, 255, 255; orange = 255, 128, 0; white = 255, 255, 255; black = 1, 1, 1). Colors were chosen without replacement within each hemifield, and colors could be repeated across, but not within, hemifields. After the retention interval, the response screen then appeared. The response screen consisted of one object on each side of the screen. Participants had to report whether the color of the object on the cued side changed colors. On a subset of trials, a series of colored squares appeared on the midline during the retention interval. These trials were excluded from analysis.

University of Chicago study #5. The experimental parameters replicate those from Chicago study #4, except for the following changes. In Chicago study #5 experiment 1, the memory array remained on the screen throughout the delay on a subset of trials. These trials were excluded from analysis. However, because the memory array remained on the screen throughout the delay, the response screen was different than all of the other experiments. On

each side of the screen, there was one object with two colors on each side of the screen (Tsubomi et al., 2013), and participants had to report which of the two presented colors matched the original memory item. In Chicago study #5 experiment 2, the memory array consisted of two or four colored squares on each side of the screen, and following the memory array, the screen remained blank for 1,650 ms.

University of Chicago study #6. This study included two experiments. Stimuli in all experiments were presented on a 24-in. LCD computer screen (BenQ XL2430T; 120-Hz refresh rate) on a Dell Optiplex 9020 computer. Participants were seated with their heads on a chin rest 74 cm from the screen. Each trial began with a blank inter-trial interval (1000 ms), followed by a gray cross (600 ms). The cue was either pointed to the attended location (50% of trials) or was uninformative (50% of trials). Following the cue, the memory array appeared (200 ms), which consisted of two (50% of trials) or four (50% of trials) colored target squares that appeared either to the left, right, above or below fixation. Four colored interrupter circles also appeared in a position adjacent to the target squares. The display was visually balanced with four gray (RGB value: 128, 128, 128) circles across from both the target squares and the interrupter circles, which matched the average luminance of the colors. Participants were told to remember the colors of the squares over the delay and ignore the circles. Squares had a side length of 0.9° visual angle and circles had a diameter of 1.0° (squares and circles covered the same area, viz., 3600 pixels). Following the memory array, the screen remained blank for 1000 ms. After this, the response screen appeared, which consisted of one colored square. Participants had to determine whether the colored square was the same color as the original-colored square in that location. Colors for the target squares and interrupter circles were selected randomly from a set

of nine possible colors (RGB values: red = 255, 0, 0; green = 0, 255, 0; blue = 0, 0, 255; yellow = 255, 255, 0; pink = 255, 0, 255; cyan = 0, 255, 255; purple = 128, 0, 255; dark green = 4, 150, 60; orange = 255, 128, 0). Colors were chosen without replacement.

EEG acquisition and artifact rejection.

University of Oregon. EEG was recorded from 22 standard electrodes sites in an elastic cap (ElectroCap International, Eaton, OH) spanning the scalp, including International 10/20 sites F3, Fz, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, T5, T6, O1, and O2, along with nonstandard sites OL, OR, PO3, PO4, and POz. Two additional electrodes were positioned on the left and right mastoids. All sites were recoded with a left-mastoid reference, and the data were re-referenced offline to the algebraic average of the left and right mastoids. To detect blinks, vertical electrooculogram (EOG) was recorded from an electrode mounted beneath the left eye and referenced to the left mastoid. The EEG and EOG signals were amplified with an SA Instrumentation amplifier (Fife, Scotland) with a bandpass of 0.01–80 Hz and were digitized at 250 Hz in Labview 6.1 running on a PC. Offline, data were low pass filtered at 50 Hz to eliminate 60 Hz noise from the CRT monitor. Eye movements (>1°), blinks, blocking, drift, and muscle artifacts were detected by applying automatic criteria. This pipeline differs from the pipeline that was used in the original paper (Unsworth et al., 2015), which is why the current analyses include more participants. A sliding-window step function was used to check for eye movements in the EOG channels. We used a split-half sliding-window approach (window size = 100 ms, step size = 50 ms, vertical threshold = 75 μV, horizontal threshold = 15 μV).

University of Chicago. EEG was recorded from 30 active Ag/AgCl electrodes (actiCHamp, Brain Products, Munich, Germany) mounted in an elastic cap positioned according to the international 10-20 system (Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz). Two additional electrodes were affixed with stickers to the left and right mastoids, and a ground electrode was placed in the elastic cap at position Fpz. Data were referenced online to the right mastoid. For Chicago studies #1-5, data were re-referenced offline to the algebraic average of the left and right mastoids, and incoming data were filtered (low cutoff = .01 Hz, high cutoff = 80 Hz; slope from low to high cutoff = 12 dB/octave) and recorded with a 500-Hz sampling rate. For Chicago study #6, data were re-referenced offline to the average of all electrodes, and incoming data were filtered (low cutoff=.01 Hz, high cutoff=250 Hz; slope from low to high cutoff = 12 dB/octave) and recorded with a 1000-Hz sampling rate. For all datasets, impedance values were kept below 10 kΩ. Eye movements and blinks were monitored using electrooculogram (EOG) activity and eye tracking. EOG data were collected with five passive Ag/AgCl electrodes (two vertical EOG electrodes placed above and below the right eye, two horizontal EOG electrodes placed ~1 cm from the outer canthi, and one ground electrode placed on the left cheek). Eye tracking data was collected using a desk-mounted EyeLink 1000 Plus eye-tracking camera (SR Research, Ontario, Canada) sampling at 1,000 Hz.

For Chicago studies #1-5, eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic criteria. After automatic detection, trials were manually inspected to confirm that detection thresholds were working as expected. For the automatic eye movement detection pipeline, a sliding-window step function was used to check for eye movements in the horizontal EOG (HEOG) and the eye-tracking gaze coordinates. For

HEOG rejection, we used a split-half sliding-window approach (window size = 100 ms, step size = 10 ms, threshold = 20 μV). HEOG rejection was only used if the eye-tracking data were bad for that trial epoch. We slid a 100-ms time window in steps of 10 ms from the beginning to the end of the trial. If the change in voltage from the first half to the second half of the window was greater than 20 μV, it was marked as an eye movement and rejected. For eye-tracking rejection, we applied a sliding-window analysis to the x-gaze coordinates and y-gaze coordinates (window size = 100 ms, step size = 10 ms, threshold = 0.5° of visual angle). We additionally used a sliding-window step function to check for blinks in the vertical EOG (window size = 80 ms, step size = 10 ms, threshold = 30 μV). We checked the eye tracking data for trial segments with missing data points (no position data are recorded when the eye is closed). We checked for drift (e.g., skin potentials) by comparing the absolute change in voltage from the first quarter of the trial to the last quarter of the trial. If the change in voltage exceeded 100 μV, the trial was rejected for drift. In addition to slow drift, we checked for sudden step-like changes in voltage with a sliding window (window size = 100 ms, step size = 10 ms, threshold = 100 μV). We excluded trials for muscle artifacts if any electrode had peak-to-peak amplitude greater than 200 μV within a 15-ms time window. We excluded trials for blocking if any electrode had at least 30 time points in any given 200-ms time window that were within 1μV of each other.

For Chicago study #6, eye movements, blinks, blocking, drift, and muscle artifacts were detected by applying automatic criteria only. To identify eye-related artifacts, eye-tracking data were first baselined identically to EEG data (i.e., subtraction of the mean amplitude of x and y coordinates for the time from −200 to 0 ms). Then, the Euclidian distance from the fixation cross was calculated from baselined data. Saccades were identified with a step criterion of 0.6° (comparing the mean position in the first half of a 50 ms window with the mean position in the

second half of a 50-msec window; window moved in 20 ms steps). Drifts were identified by eye-tracking data indicating a distance from the fixation of >1°. Both eyes had to indicate an eye-related artifact for a trial to be excluded from analysis. In addition, trials in which any EEG channel showed a voltage of more than 100 μV or less than −100 μV were rejected.

Behavioral analysis. Working memory capacity (*K score*) was used to measure task performance (Cowan, 2001; Pashler, 1988). Capacity was calculated with the formula $K = S(H - F)$, where K is working memory capacity, S is the size of the array, H is the observed hit rate, and F is the false alarm rate. *K scores* were calculated with different set sizes within the Oregon- and Chicago-site data. Within the Oregon-site data, *K scores* were calculated from a separate working memory task collected outside the EEG booth that that had both set size 2 and 6 for all participants. For the Chicago-site participants, *K score* was calculated using different set sizes, depending on which experiment(s) the participant completed. Chicago-site set sizes included 2, 3, 4, and 6 items. In the Chicago sample, there was no significant relationship between the number of trials that a participant completed and *K score*: r=0.13, p=0.10, *mse*=1.65e6.

Calculation of inter-electrode correlations using EEG.

Electrode organization. We analyzed EEG data that was collected while participants performed a lateralized change detection task. In this type of task, participants attend and maintain information on either the left or right side of the screen on any given trial. This results in lateralized neural activity that is contralateral to the remembered items (Vogel & Machizawa, 2004b). For example, if stimuli are presented on the left side of the screen, this lateralized neural activity would be present in electrodes O2, PO8, PO4, etc., whereas if stimuli were presented on

the right side of the screen, it would be present in electrodes O1, PO7, PO3. We accounted for this lateralized neural activity by aligning neural activity based on which side of the screen participants were attending on each trial. For example, activity for electrode number 10 in the matrix included data from PO8 on "attend-left" trials and data from PO7 on "attend- right" trials. Central electrodes (Fz, Cz, and Pz) were unaffected by this organization. This method of alignment is similar to how CDA analyses account for the contralateral organization of the visual system (Vogel & Machizawa, 2004b).

Inter-electrode correlations of ERP activity. The Chicago and Oregon datasets had different numbers of electrodes. Therefore, in our analyses we only included the 17 electrodes that overlapped between the two datasets: F3, Fz, F4, C3, Cz, C4, P7, P3, Pz, P4, P8, PO8, PO4, PO3, PO7, O1, and O2.

Previous fMRI connectome-based models have concatenated task data across trials or runs, and then correlated these concatenated time series across all pairwise nodes (here, electrodes). Because EEG data are noisy when concatenated across trials, we instead averaged the raw amplitude at each of these 17 electrodes across all trials from timepoints 0 to 1000 ms. This is analogous to calculating one ERP time course for each electrode. Timepoint zero corresponded to the onset of the memory array, and timepoint 1000 was the end of the shortest retention interval (1000 ms). We analyzed data from the same time window for all experiments to match the number of timepoints included in the analysis across experiments. We computed the Pearson correlation of this trial-averaged EEG activity for all pairwise electrodes for each participant separately. For each participant, this resulted in a 17 x 17 matrix of the correlation between the time course of each electrode to each other electrode. We Fisher z-transformed

correlation coefficients and submitted the resulting ERP correlation matrices to the analyses described below.

Statistical analyses. We tested all analyses within the Oregon dataset, whenever this was possible. We then replicated these analyses and externally validated predictive models in the compiled Chicago dataset.

EEG fingerprinting. This analysis investigates whether individuals' inter-electrode correlation patterns are unique and stable enough to distinguish them from a group. These methods were adopted from a previous paper that investigated functional connectome fingerprinting in fMRI (Finn et al., 2015).

To identify individuals in the Oregon dataset, we compared participants' vectorized shape-task and color-task inter-electrode correlation matrices. Specifically, for each individual, we correlated their "target" color-task vector of inter-electrode correlations from with a "database" of all shape-task vector of inter-electrode correlations. An individual was considered accurately identified if the maximum correlation was with their own shape-task data. We repeated this analysis using shape-task vectors as the "targets" and color-task vectors as the "database". Finally, we characterized EEG fingerprinting accuracy as the number of correct identifications divided by the number of participants*2.

For the Chicago study, we analyzed data from the subset of 45 individuals who participated in more than one experiment. For each person in this sample, we first calculated and vectorized an inter-electrode correlation matrix from each experiment in which they participated separately. Next, we selected one of these vectors to serve as the "target". We compared this

target to a database, which included 19 individuals' vectors of inter-electrode correlation from another task in which the target individual also participated (including the target individual). We chose 19 because this was the minimum number of participants in any individual study, and we wanted to equate chance levels across the experiments. To predict subject identity, we computed the similarity between the target vector of inter-electrode correlations and each database vector, and the predicted identity was that with the maximal similarity score. Similarity was defined as the Pearson correlation between the vectors.

To ask whether the specific inter-electrode correlations ("edges") that predict behavior contribute to identification more than expected by chance, we ran these analyses including only those edges that significantly predicted behavior (see ***EEG-based predictive modeling*** below). To determine whether these results were due to down sampling the number of edges included in the ERP correlation matrix, we compared the results from the significant edge analysis to an analysis with the same number of randomly selected edges.

To assess the statistical significance of identification accuracy, we performed non-parametric permutation tests. We ran the same analyses as above, except we shuffled the subject labels in each iteration, so that they would randomly align with the ERP correlation data. This shuffling of labels was repeated 10,000 times.

EEG-based predictive modeling. Prediction methods were adopted from previous work using fMRI functional connectivity to predict behavior (Rosenberg et al., 2016; Shen et al., 2017). We first separated data into training and testing sets. For internally validated models, we used 5-fold cross-validation and permutation testing to determine whether the model

significantly predicted behavior. For external validation analyses, we trained models using data collected in Oregon and tested them using data collected in Chicago and vice versa.

To identify the edges that were significantly related to behavior in the training sample, we correlated each value in the matrix of inter-electrode correlations to K score using Spearman's correlation. This approach replicated the connectome-based predictive modeling pipeline applied to fMRI data (Shen et al., 2017) and allowed us to compare the anatomy of the inter-electrode correlation feature sets ("networks") positively and negatively correlated with behavior. For the internally validated models and the externally validated model trained on the Chicago dataset, we selected the top 5% of edges that most strongly predicted behavior in the positive and negative directions. For the externally validated model that was trained on all of the Oregon data, we identified the 5% of edges most positively and the 5% of edges most negatively correlated with working memory capacity (10% of total edges) in the color and shape task separately. We defined predictive edges as those that were included in the top 5% of edges positive and negatively predicting behavior in both tasks in order to get the most robust predictive feature set. We did not perform this overlap analysis for the externally validated model that was trained on the Chicago dataset because each task had a relatively small sample size (*n*s = 19–29), which could result in unreliable features.

Using these selected features, we then calculated single-subject summary values for all training subjects. To do this, we summed the correlation-strength values for each individual for both the positively and negatively predictive feature sets separately, and then we took the difference between them. We trained a linear model with these summary features and K scores from the training set to predict behavior from inter-electrode correlations. We then used this model to predict K score in the testing set. To do this, we calculated summary features in test set

participants, and input these summary scores into the model defined in the training sample to generate predictions for each test set subject's K score. Results are reported as p-values, r-values, and mean square error (mse).

Predictive network overlap. We calculated whether the predictive edges from the Oregon and Chicago models significantly overlapped. We determined significance of network overlap using the hypergeometric cumulative density function in Matlab (Rosenberg et al., 2016). The formula was as follows: p=1-hygecdf(x, M, K, N), where x is the number of overlapping edges, M is the total number of edges in the matrix, K is the number of edges in the Oregon network, and N is the number of edges in the Chicago network.

Relationships between predicted working memory and other cognitive abilities. Finally, we asked whether models that predicted working memory capacity across individuals also predicted other cognitive abilities. We ran these analyses only with the Oregon dataset because these participants completed behavioral tasks assessing a range of cognitive abilities in addition to the change detection tasks used to measure K. These analyses included the 138 participants who completed all cognitive tasks.

We were specifically interested in the relationship between inter-electrode correlations and gF. To investigate this relationship, for each subject in the Oregon dataset, we calculated the strength in the working memory network defined in the Chicago sample. Then, using 5-fold cross-validation, we defined a linear model relating network strength to gF. We then used this model to predict the left-out set of participants' gF scores. Previous research has shown that gF and K score are highly correlated (Conway et al., 2002; Unsworth et al., 2015). Given this

relationship, we calculated a theoretical ceiling of these models' performance as the correlation between gF and K scores.

As an exploratory analysis, we also investigated the relationship between predicted K scores and performance on all other behavioral tasks (**Table 1**). To do this, we computed the Pearson correlation between 1) observed K score and all of the other tasks (**Table 1**, column 2) and 2) predicted K score and all of the other tasks (**Table 1**, column 3). Then, we correlated these two columns of data to determine whether there was a significant relationship between performance on all tasks and the measured and predicted K scores. We did this separately for the color and the shape tasks.

Results

EEG fingerprinting

Are patterns of inter-electrode correlations unique across individuals? Many EEG analyses implicitly treat potential individual differences as noise, averaging results across the group or comparing average results from two groups (e.g., individuals with high vs. low working memory capacity; patients vs. controls). Here, we asked whether individuals have both *stable* and *unique* trial-evoked patterns of inter-electrode correlations that can reliability distinguish them from others. To test this possibility, we applied "EEG fingerprinting", an approach developed using fMRI functional connectivity data (Finn et al., 2015), to two independent EEG datasets. Data were collected as participants performed variants of a lateralized change detection task at the University of Oregon (n = 171) and the University of Chicago (n = 165; see Methods for details). Fingerprinting analyses were restricted to individuals who participated in at least two separate tasks at one of the sites (Oregon n=171; Chicago n=45).

University of Oregon sample. Participants in the University of Oregon sample completed two lateralized change detection tasks in which they were instructed to remember items' shape or color as part of a single study. EEG fingerprinting analyses revealed that individuals had stable and unique patterns inter-electrode of correlations across these two tasks that could reliably dissociate them from other individuals: color: shape task accuracy = 82%, $p<0.0001$; shape: color task accuracy = 81%, $p<0.0001$ (chance = 1/171, or .58%). In other words, individuals' patterns of inter-electrode correlations were distinct from the group and stable across tasks. Furthermore, individuals' patterns of inter-electrode correlations looked most like themselves, regardless of task context (within-subject, across-task average matrix correlation=0.9992+/–0.00058; across-subject, within-task average matrix correlation =0.9941+/–0.0039; $p<0.0001$).

To account for potential differences in skull thickness across individuals that could affect inter-electrode correlations and result in an overestimation of identification accuracy, we applied a Laplacian transformation to the time series data and then re-ran identification analyses. Even after accounting for potential differences in volume conduction, identification accuracy remained high: color: shape task accuracy = 77.78%, $p<0.001$; shape: color task accuracy = 80.12%, $p<0.001$ (chance = 1/171, or .58%). This aligns with previous research that has found that differences in brain activity contribute much more variance to surface EEG than do variations in skull thickness (Hagemann et al., 2008).

Are patterns of inter-electrode correlations unique in their ability to identify individuals, or are trial-averaged amplitude patterns stable and distinguishable across individuals as well? To address this question, we performed fingerprinting analyses using trial-averaged amplitude at

each electrode, instead of inter-electrode correlations of trial-averaged amplitude. Results revealed that trial-evoked amplitude significantly identified individuals: color: shape task accuracy = 51%, p<0.0001; shape: color task accuracy = 51%, p<0.0001 (chance = 1/171, or .58%). However, identification accuracy was lower than that achieved with patterns of inter-electrode correlations (color: shape task accuracy = 82%, p<0.0001; shape: color task accuracy = 81%, p<0.0001).

University of Chicago sample. The full University of Chicago sample included data from twelve experiments collected as part of six independent studies. Data from all studies were collected on different days in different sessions. For this analysis, we only consider data from the subset of individuals who participated in multiple experiments (see Methods for detail). Replicating results from the Oregon sample, individuals in the Chicago dataset showed stable and unique patterns of inter-electrode correlations, which could be used to reliably distinguish them from other individuals: task *A*:*B accuracy*=29%, *p*<0.0001; task *B*:*A accuracy*=34%, *p*<0.0001 (chance = 1/19, or 5.26%). Additionally, individuals' patterns of inter-electrode correlations looked most like themselves, regardless of task context (across-task matrices average=0.9955+/–0.0046; within-task matrices average=0.9923+/–0.0063; p<0.0001). Once again, identification accuracy was not driven by differences in skull thickness: Laplacian transformed identification of task *A:B accuracy*=37.62%, *p*<0.001; task *B:A accuracy*=36.00%, *p*<0.001 (chance = 1/19, or 5.26%). Additionally, trial-evoked amplitude significantly identified individuals: task *A:B accuracy*=17%, p<0.0001; task *B:A accuracy*=18%, p<0.0001 (chance = 1/19, or 5.26%). However, this identification accuracy was lower than that achieved with

patterns of inter-electrode correlations: task *A*:*B* *accuracy*=29%, $p<0.0001$; task *B*:*A* *accuracy*=34%, $p<0.0001$ (chance = 1/19, or 5.26%).

In both the Chicago and Oregon datasets, we identified individuals based on their unique pattern of inter-electrode correlations. These results suggest that individuals have unique and robust inter-electrode correlation patterns that differentiate them from others. To test whether individual differences in these patterns reflect individual differences in cognitive abilities and behavior, we next asked whether we could use these patterns to predict a central cognitive ability: working memory capacity.

Predicting working memory: Within-site validation

To determine whether patterns of inter-electrode correlations during lateralized change detection tasks predicted working memory capacity in novel individuals, we trained and tested EEG-based predictive models (Rosenberg et al., 2016; Shen et al., 2017) using balanced 5-fold cross-validation. For each participant, we calculated working memory capacity (K score) based on their change detection task performance (Cowan, 2001; Pashler, 1988). We calculated inter-electrode correlations for each participant collapsed across all tasks, as described in the Methods. As a comparison, we also predicted working memory capacity using amplitude, rather than inter-electrode correlations of amplitude.

University of Oregon sample. We averaged each participant's pattern of inter-electrode correlations from the color and the shape change detection tasks and built models to predict working memory capacity measured in a completely independent task. Models based on inter-

electrode correlations significantly predicted novel individuals' working memory capacity: median correlation between predicted and observed K score r=0.25, p<0.0085, *mse*=0.24; Figure 5-1. Control analyses revealed that models based on EEG signal amplitude, rather than inter-electrode correlations of amplitude, did not predict capacity: r=0.02, p=0.43, *mse*=0.26.

Next, we built predictive models for the color and shape tasks separately. We found that observed and predicted K scores were significantly correlated in the shape task, but not the color task: median shape task, r=0.26, p<0.0045, *mse*=0.42; color task, r=0.13, p<0.08, *mse*=0.25.

We next investigated whether models generalized across tasks by applying models trained in the color task to shape task data from held-out individuals and vice versa. Models trained on color task data significantly predicted K scores from shape-task data: median correlation between predicted and observed K score r=0.26, p<0.0045, *mse*=0.45. Models trained on data from the shape task, however, did not significantly predict K scores from color-task data: r=0.13, p<0.08, *mse*=0.42.

The above models were built using patterns of inter-electrode correlations calculated from trial-averaged ERPs from electrodes aligned based on the side of the screen that participants attended on each trial as described in the *Electrode organization* section of the Methods. When electrodes were not aligned in this way, models did not predict working memory capacity: median r=-0.04, p=0.65, *mse*=0.27. Similarly, models trained on correlation patterns calculated from EEG data concatenated rather than averaged across trials did not predict capacity: median r=0.02, p=0.39, *mse*=0.27. Critically, these approaches were tested in the Oregon dataset only before successful models were replicated and externally validated in the Chicago dataset.

University of Chicago sample. For the Chicago dataset, we collapsed data across multiple lateralized change detection experiments, which had different memoranda and sample durations. Inter-electrode correlations and K for each participant were measured using all of the studies that each participant completed. We calculated K for each participant from the EEG task data. We once again applied 5-fold cross validation to determine whether patterns of inter-electrode correlations could predict working memory capacity across individuals. Replicating findings from the Oregon sample, observed and predicted K scores were significantly correlated: median r=0.24, p<0.0044, mse=0.33; Figure 5-1. Thus, inter-electrode correlations significantly predict working memory capacity across individuals.

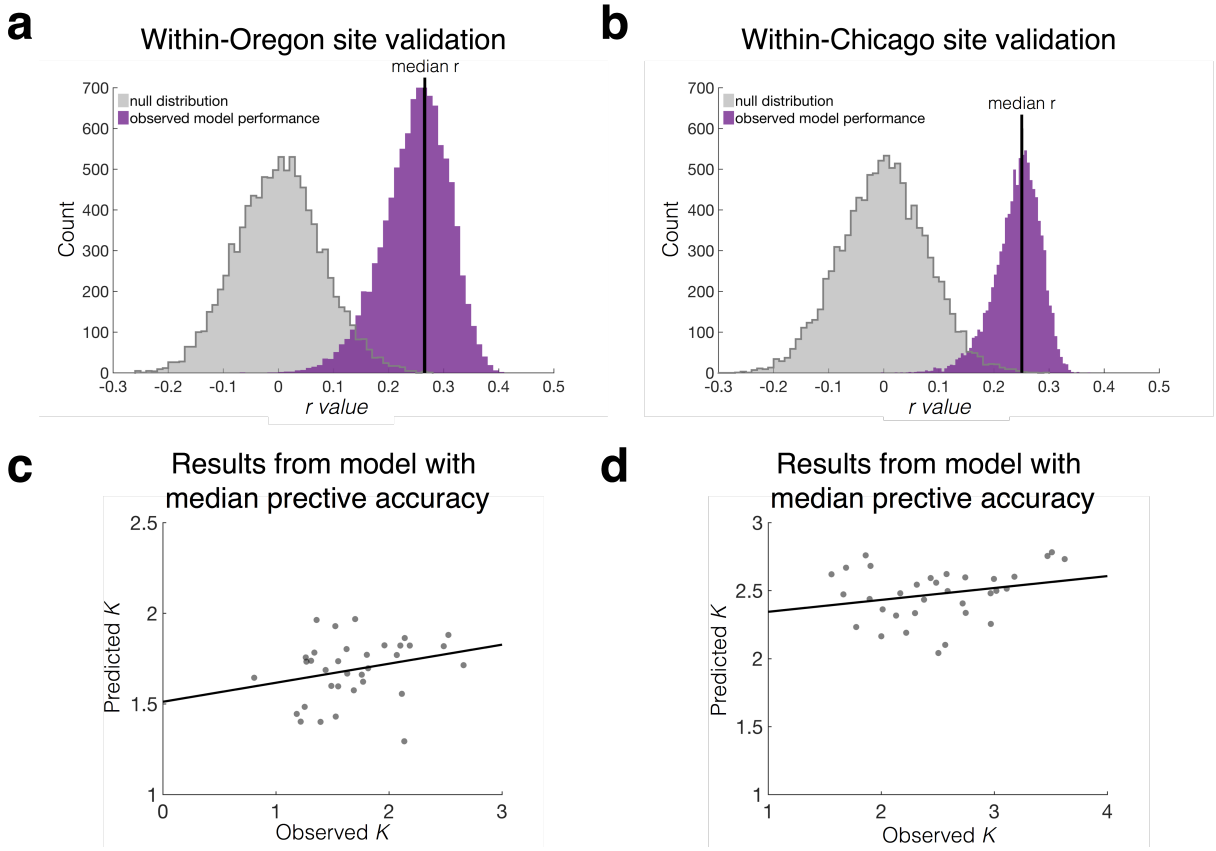

Figure 5-1. Within-site validation results.

Histogram of the correlation between observed and predicted working memory capacity (*K*) using 5-fold cross-validation over 10,000 true model iterations (purple) and 10,000 null model

(Figure 5-1 continued) permutations (gray) for (**a**) the model trained within the Oregon-site data and (**b**) the model trained within the Chicago-site data. The vertical black lines represent the iteration and fold with the median *r* value. Scatter plot of the correlation between observed and predicted K scores from the iteration and fold with the median *r* value for (**c**) the model trained within the Oregon-site data and (**d**) the model trained within the Chicago-site data. Gray dots represent individuals observed and predicted K scores from one fold. The black line is the best fit line.

Predicting working memory: Across-site validation

Although these results provide the first evidence that models based on EEG data predict working memory capacity across individuals, there are many reasons that internal (i.e., within-dataset) validation may overestimate effect sizes, including idiosyncrasies in task context, EEG systems, and participant populations. Therefore, an even more powerful demonstration of the robustness of predictive models is to externally validate them—that is, to apply them to data from a completely independent sample. External validation allows us to better approximate a model's population-level generalizability and better understand its predictive boundaries.

To test the cross-dataset generalizability of models predicting working memory capacity, we trained models on the full Oregon sample and applied them to the full Chicago sample and vice versa. Importantly, these two datasets were collected by different experimenters in different locations using different EEG systems.

Predicting Chicago K scores from Oregon model. To generalize the Oregon working memory model to the Chicago dataset, we trained a model using the Oregon dataset. Predictions from this model were significantly correlated with observed K scores, demonstrating that the model trained on the Oregon dataset significantly predicted working memory capacity in the Chicago dataset: r=0.24, p=0.002, *mse*=0.95; Figure 5-2. Cross-site prediction performance was robust to variations in the feature-selection threshold applied during model building. Thresholds

ranging from the top 5% to the top 20% of edges positively and negatively correlated with behavior resulted in model performance values between $r = 0.19 – 0.31$, $p < 0.01$, $mse = 0.95 – 1.08$.

Predicting Oregon K scores from Chicago model. We next applied a model defined using Chicago data to predict working memory in the Oregon sample. Predictions from this model were significantly correlated with observed K scores, demonstrating that the model trained on the Chicago dataset significantly predicted working memory performance in the Oregon dataset: r=0.28, p=0.0003, mse=1.06; Figure 5-2. Again, cross-site prediction performance was robust to variations in feature-selection threshold. Thresholds ranging from the top 5% to the top 20% of edges positively and negatively correlated with behavior resulted in model performance values between $r = 0.18 – 0.28$, $p < 0.02$, $mse = 0.95 – 1.08$.

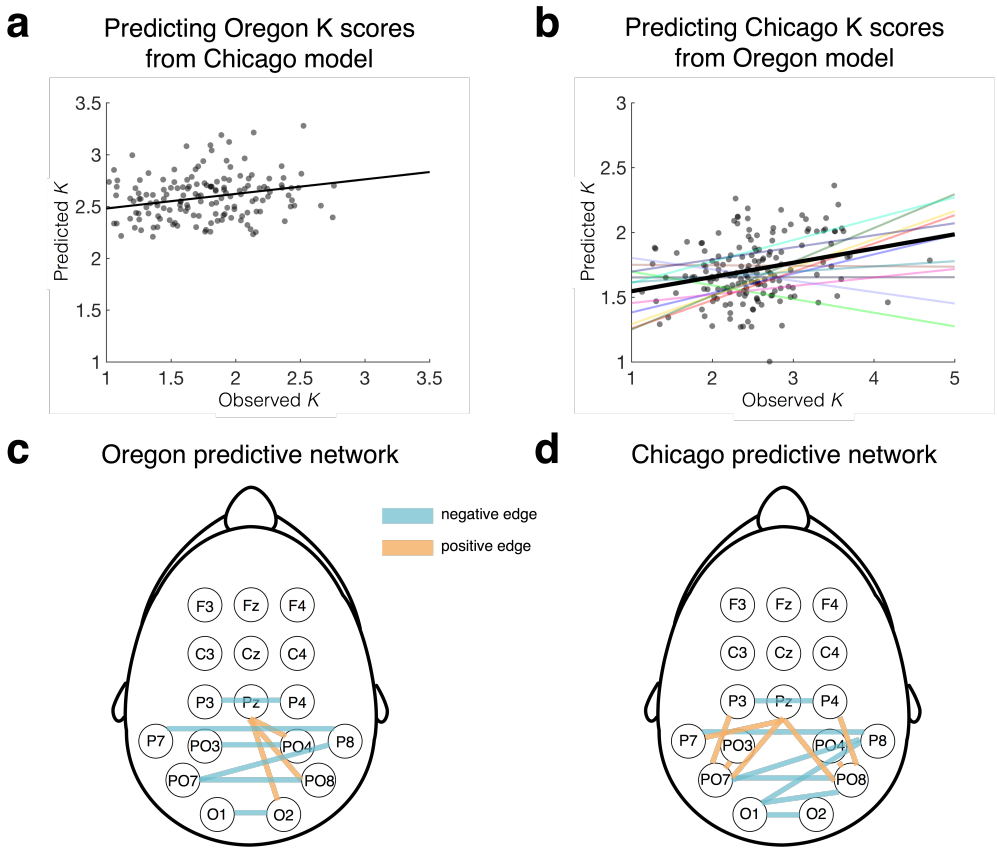

Figure 5-2. Across-site validation results.

a Scatter plot of the correlation of observed and predicted working memory capacity (*K*) from a model trained on all of the Chicago data and tested on all of the Oregon data. Gray dots represent individuals observed and predicted K scores. The black line is the best fit line. **b** Scatter plot of the correlation of observed and predicted working memory capacity (*K*) from a model trained on all of the Oregon data and tested on all of the Chicago data. Individual colored lines represent the correlation between observed and predicted working memory capacity of each experiment plotted separately and demonstrate that it is not the case that a small subset of experiments drives these results. The black line represents the fit of all individual participants. **c** Predictive edges (i.e., inter-electrode correlations) that were included in the model trained on the Oregon dataset. Edges that positively predicted working memory capacity are depicted in orange, and those edges that negatively predicted working memory capacity are depicted in blue. **d** Predictive edges that were included in the model trained on the Chicago dataset.

With this across-site validation, we demonstrate that models predicting working memory capacity not only generalized across EEG session and stimuli, but also generalized across data acquisition sites (Oregon vs. Chicago) and EEG systems (passive vs. active). This suggests that

inter-electrode correlations are a robust, reproducible, and generalizable predictor of working memory capacity.

Predictive network anatomy

Overlap of Oregon and Chicago predictive networks. Next, we asked whether the two externally validated models predicted working memory from a set of common edges, or network. We found that indeed there was significant overlap between the predictive edges in the Chicago and Oregon networks: 6 overlapping edges (26% of the edges in both networks; 5 edges negatively predicting behavior in both samples and one edge positively predicting behavior in both samples), p=7.68e-7. This suggests that a common set of inter-electrode correlations predicts working memory capacity in both samples.

Network anatomy. The edges that overlapped between the Oregon and Chicago networks involved posterior and occipital electrodes: O1-O2, PO8-PO7, P7-P8, PO7-P8, P4-P3, PO8-Pz. Most of these edges were between cross-hemisphere electrodes, which could be due to the lateralized nature of the task. Interestingly, this predictive network did not include any frontal edges. One possible explanation for the lack of frontal edges could be that frontal edges tend to be noisier than posterior edges, potentially even after artifact rejection. Given this, we sought to determine whether frontal edges were less reliable than other edges, and if that reliability was correlated with predictability across all edges. To do this, we calculated split-half reliability of each electrode. Then, we compared each electrodes' reliability score with their predictability—as measured by the correlation between edge strength and behavior, averaged for each electrode. We found that there was not a significant relationship between electrode reliability and

predictability: Oregon, $r = –0.13$, $p = 0.62$; Chicago, $r = 0.40$, $p = 0.12$. Therefore, the lack of frontal edges in our working memory models is unlikely to be due to an increase in noise in those edges relative to other edges.

Interestingly, predictive networks in our models include electrodes that are typically used to calculate the CDA. This large degree of overlap between our models' predictive networks and CDA electrodes raises some questions about whether our models are independent from the CDA. To answer this question, we ran the same prediction analyses as we did with the above EEG models, but instead of using inter-electrode correlations, we used CDA amplitude to predict working memory capacity. We calculated the CDA by taking the difference in amplitude between contralateral and ipsilateral posterior/occipital electrodes (P7/P8, P3/P4, PO7/PO8, PO3/PO4, O1/O2) from 0-1000 ms (the time range included in our analyses). We then trained a linear model to predict working memory capacity from the CDA using the full Oregon-site dataset and tested it on the full Chicago-site dataset (and vice versa). The model trained on the Oregon-site dataset did not significantly predict working memory capacity: r=0.04, p=0.61. Surprisingly, the model trained on the Chicago-site dataset *negatively* predicted working memory capacity: r=–0.32, p=1.74e–5. This negative prediction is difficult to interpret, especially because this model predicted K score values that ranged from 2.50 to 2.52.

Past work has shown that CDA amplitude in the Oregon-site dataset is significantly correlated with working memory capacity (Unsworth et al., 2015). However, in the Chicago-site data, the correlation between CDA amplitude and working memory capacity was not significant: $r = –0.04$, $p = 0.61$. One possible reason for the lack of a significant correlation in this dataset is that the set sizes used to calculate K scores are smaller than set sizes typically used to observe this correlation. Smaller, variable set sizes could affect estimates of working memory capacity

and, thus, the relationship between K scores and CDA amplitude. Nevertheless, the inter-electrode correlation approach seems to be more sensitive than the CDA to the relationship between neural activity and behavior because the inter-electrode correlation approach was able to predict behavior in both datasets despite being calculated using a more limited behavioral range in the Chicago-site dataset.

To further investigate whether CDA explains unique variance to inter-electrode correlations, we trained models to predict working memory capacity from both CDA and inter-electrode correlations, combined. These models significantly predicted behavior: train Oregon, test Chicago r=0.25, p=0.002; train Chicago, test Oregon r=0.27, p=0.0004. However, the CDA coefficient was only significant in the model trained on the Oregon data: train Oregon CDA coefficient p=0.02; train Chicago CDA coefficient p=0.61. These results provide further evidence that our inter-electrode correlation models do not simply track the CDA. Instead, they predict behavior from neural signals that are distinct from the CDA.

EEG fingerprinting using predictive networks.

Are the EEG networks that predict working memory capacity across individuals reliable enough to distinguish individuals from a group? To ask this question, we applied the same analysis described in the *EEG fingerprinting* section above. However, instead of using the whole-scalp pattern of inter-electrode correlations, we only included the inter-electrode correlations that significantly predicted behavior across individuals. This network only included those edges that significantly predicted behavior in both the Chicago and Oregon models. We then compared the results from the predictive network to those from the whole-scalp network.

University of Oregon sample. Edges predicting working memory were sufficient to significantly identify individuals: color: shape task accuracy=40%, $p<0.0001$; shape: color task accuracy=39%, $p<0.0001$ (chance = 1/171). In other words, the EEG networks that predicted working memory capacity across individuals also reliably distinguished individuals. However, identification accuracy using only predictive edges was lower than when we included the whole-scalp network. To determine whether this reduction in identification accuracy was due to downsampling the number of edges, we compared identification accuracy of predictive edges to an equal number of randomly selected edges that were not predictive of working memory performance. Identification accuracy using a random subset of edges also significantly identified individuals: color: shape task accuracy=39%, $p<0.0001$; shape: color task accuracy=33%, $p<0.0001$ (chance = 1/171). There was not a significant difference between identification accuracy using the predictive edges and the random edges: color: shape $p<0.167$; shape: color $p<0.272$. These results suggest that edges that significantly predict working memory capacity are not necessarily better at identifying individuals than a random subset of edges.

University of Chicago sample. Replicating results from the Oregon sample, identification accuracy using only predictive edges was successful: task *A*:*B* accuracy=26%, $p<0.0001$; task *B*:*A* accuracy=23%, $p<0.0001$ (chance = 1/19). This suggests that the network that predicted working memory capacity across individuals was also stable across tasks and was able to distinguish individuals from a group. Once again, identification accuracy using only predictive edges was worse than when we included the whole-scalp network. To determine if this decrease in identification accuracy was due to downsampling the number of edges, we calculated identification accuracy of a random subset of edges. We found that these random edges also

significantly identified individuals: task *A*:*B accuracy*=13%, $p<0.0001$; *B*:*A accuracy*=13% $p<0.0001$ (chance = 1/19). However, unlike in the Oregon dataset, the predictive edges more accurately identified individuals than the random edges: task *A*:*B* $p<0.0001$; task *B*:*A* $p<0.0001$. These results suggest that edges that significantly predict working memory capacity may more accurately identify individuals than a random subset of edges. Given the inconsistency of this result across the Chicago and Oregon samples, further research is needed to characterize whether an individual's most identifiable edges are the same edges that best predict their behavior.

Relationship between inter-electrode correlations and other cognitive abilities

Our results demonstrate that patterns of inter-electrode correlations observed during a working memory task are a robust and reliable predictor of working memory capacity. Do patterns of inter-electrode correlations observed in this context predict other cognitive abilities? In the Oregon dataset, participants performed a series of cognitive tasks outside the EEG booth (see Methods section for details). Therefore, we used this dataset to investigate the relationship between patterns of inter-electrode correlations and other cognitive abilities.

Predicting fluid intelligence. Although we have shown that the pattern of inter-electrode correlations is a robust predictor of working memory capacity, one concern is that these correlations are not predicting working memory *per se*, but instead some other aspect of function that could support performance in the working memory task. For instance, inter-electrode correlations might reflect the degree of response bias in an individual, which could in turn have an impact on change detection performance. To address this possibility, we tested whether networks that predict working memory capacity also predict general fluid intelligence (gF), a

closely related ability (Conway et al., 2002; Engle et al., 1999; Fukuda et al., 2010). If we are able to predict *gF* based on our EEG networks, this would provide converging evidence that inter-electrode correlations are predicting a purer working memory capacity construct because that which is predicted is shared with a very different task that is known to rely on working memory.

To this end, we calculated gF scores for each participant in the Oregon sample from a combination of performance on three tasks: Raven advanced progressive matrices, number series, and Cattell's culture fair test, as reported in previous work (Unsworth et al., 2015). To determine whether the same EEG networks that predict working memory capacity also predict gF in novel individuals, for each subject in the Oregon dataset, we calculated the strength in the working memory network defined in the Chicago sample. Next, using 5-fold cross-validation, we defined a linear model relating these working memory network strength values to gF. We then used this model to predict the left-out set of participants' gF scores. Models significantly predicted novel individuals' gF scores: r=0.29, p<7.00e–4, *mse*=6.87; Figure 5-3. These results suggest that EEG networks tap into the same aspects of working memory capacity that are critical for fluid intelligence, alleviating concerns that EEG models were instead tapping into some other idiosyncratic factor influencing performance on the change detection task used to measure working memory ability in our study. Although the EEG data used here were collected as participants performed a working memory task, our EEG models generalized to predict fluid intelligence. Of note, this modeling approach is similar to correlating predicted K and observed gF, except that the linear model has a gF-specific coefficient that changes in every fold. Both observed and predicted K scores are correlated with gF (observed K: r=0.41, p=5.98e–7; predicted K: r=.18, p=.038), suggesting that retaining the model is not critical.

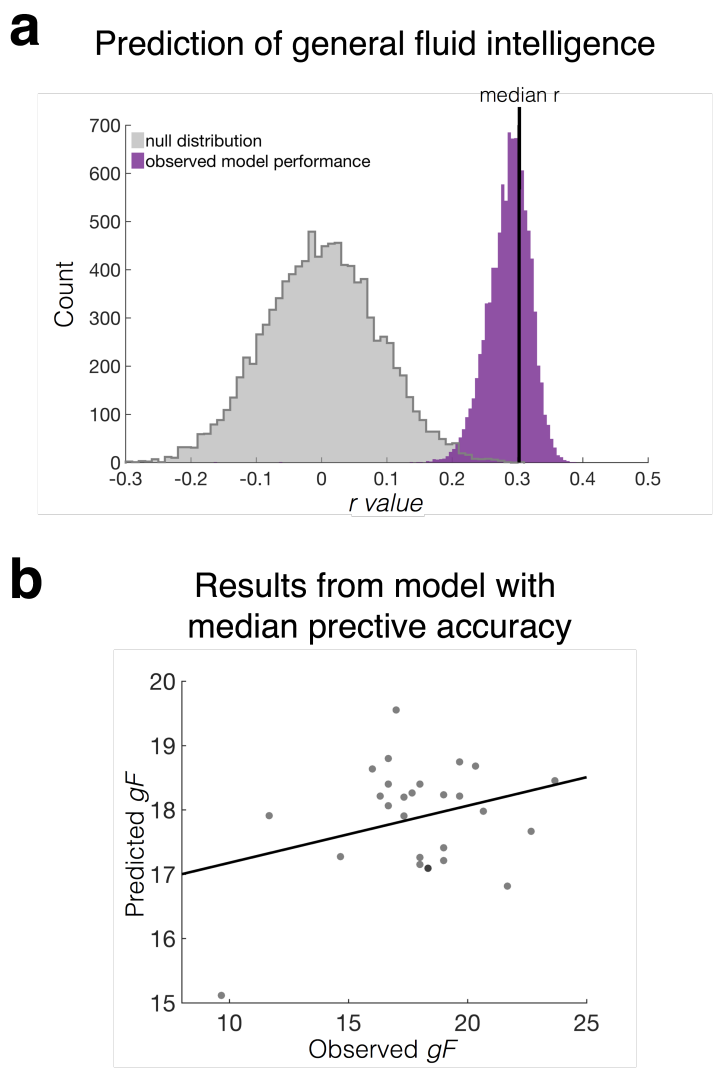

Figure 5-3. Prediction of general fluid intelligence.

a Histogram of the correlation between observed and predicted general fluid intelligence (*gF*) using 5-fold cross-validation over 10,000 true model iterations (purple) and 10,000 null model permutations (gray). The vertical black line represents the iteration and fold with the median *r* value. Models were trained and tested within the Oregon dataset using the predictive working memory edges that were included in both the Oregon and Chicago working memory models. **b** Scatter plot of the correlation between observed *gF* and predicted *gF* from the iteration and fold with the median *r* and *p* values. Gray dots represent individuals observed and predicted *gF*. The black line is the best fit line.

Correlations between predicted working memory capacity and other cognitive abilities.

We next characterized how predicted working memory capacity from our models related to fluid intelligence as well as other cognitive abilities measured in the Oregon dataset, including

attentional control and long-term memory. To investigate this, we performed a behavioral cross-correlation analysis. First, we correlated observed K scores with all other observed behavioral scores. Next, we correlated predicted K scores and all other observed behavioral scores. Correlating these two columns of data revealed that the relationship between *predicted* working memory capacity and other cognitive abilities mirrors the relationship between *observed* working memory capacity and other cognitive abilities: r=0.90, p=7.17e−10. In other words, model predictions capture variance in working memory that is shared with related cognitive abilities, rather than variance that is specific to the working memory task used in the EEG studies.

Discussion

Here, we demonstrate that inter-electrode correlations measured using EEG are idiosyncratic to each person and can be used to identify individuals from a group. These correlation patterns are unique and stable over time, just like a fingerprint. Furthermore, individual differences in these patterns are cognitively meaningful, predicting general fluid intelligence across individuals and working memory capacity across independent datasets. Together, these results demonstrate that individual differences in critical cognitive abilities are reflected in individuals' unique, idiosyncratic expression of shared EEG networks. Furthermore, patterns of inter-electrode correlations are a generalizable and accessible approach for predicting individual differences in other abilities and behaviors from brain data.

Predicting trait-like behavior

Previous research has used EEG to track moment-to-moment fluctuations in working memory storage. For example, multivariate pattern classification techniques have used the

topography of raw EEG amplitude to decode the amount of information maintained in working memory on a single trial (Adam et al., 2020). Another EEG signal, the contralateral delay activity (CDA), scales with the number of items held in working memory. This signal's amplitude asymptotes when working memory is full and is correlated with working memory capacity (Unsworth et al., 2015). Previous research has also found that ERP components, including the P200, N200, and P300, are strongly associated with general fluid intelligence (Schubert et al., 2017). Thus, the presence of EEG signals that track working memory storage, capacity, and intelligence is well established. However, no previous work has demonstrated the ability to *predict* working memory capacity and general fluid intelligence across completely independent individuals from different datasets using direct measures of brain function, including EEG.

Brain-based models of behavior are most theoretically informative and practically useful when they generalize to novel data. However, there are many reasons why brain-based models might not generalize. For example, internally validated model results could be driven by similarities between people at a particular site or idiosyncrasies of a particular task design or experimental context. Testing models on independent datasets (i.e., external validation) is a powerful way to reduce these and other biases (Poldrack et al., 2020; Woo et al., 2017). Although significant work has emphasized the importance of external validation (Poldrack et al., 2020), it is still relatively uncommon (Woo et al., 2017). In fact, one paper found that only 9% of the neuroscience studies that they surveyed tested models on one or more independent datasets (Woo et al., 2017). Additionally, none of these studies analyzed EEG data. Here, we externally validated our EEG-based predictive models. We illustrate that our results are robust to differences in task design, experimental context, EEG system, and participant population. By

externally validating our results, we provide a powerful demonstration that our models of working memory are both robust and generalizable.

Interestingly, we were able to predict general fluid intelligence using the same EEG network that predicted working memory capacity. This suggests that the variance in working memory capacity that our predictive model explains is shared with general fluid intelligence. With our cross-correlation analysis, we also found that the relationship between predicted working memory capacity and other cognitive abilities was analogous to the relationship between observed working memory capacity and other cognitive abilities. This is another example of how our model predictions capture variance in working memory that is shared with related cognitive abilities. Overall, the EEG-based model that we identified is not idiosyncratic to a particular working memory task. Rather, it seems to more generally track trait-like cognitive abilities, including general fluid intelligence.

Sparse cognitive networks

Previous work in cognitive and network neuroscience has suggested that cognitive abilities, such as working memory and attention, emerge from interactions between dozens or hundreds of brain regions in complex, large-scale brain networks (Bressler & Menon, 2010; Rosenberg et al., 2017). Significant research, typically using fMRI, has utilized these networks to describe and predict behavior (Beaty et al., 2018; Finn et al., 2015; Rosenberg et al., 2016; Sripada et al., 2013; Yamashita et al., 2018). Interestingly, using EEG, we were able to predict working memory capacity from fewer than 15 inter-electrode correlations and fluid intelligence from only 6 correlations between occipital and parietal electrodes. Thus, although cognitive

abilities may involve the interaction between large numbers of disparate brain regions, they can also be summarized using a relatively sparse EEG network.

In light of previous work that has found robust associations between frontoparietal networks and working memory, it may be surprising that our working memory models do not include correlations between frontal electrodes as features. However, the lack of frontal electrode correlations in our models does not necessarily suggest a lack of involvement of frontoparietal networks in working memory processes. EEG's lack of spatial specificity makes it very difficult, if not impossible, to localize where the signal that we are measuring was generated. Thus, it is possible that the activity in posterior electrodes that predicted working memory capacity reflected contributions from anterior brain regions that have previously been shown to be involved in working memory. Furthermore, there are many robust EEG signals that track working memory that are measured in posterior electrodes. For example, the contralateral delay activity (CDA)—which is measured using posterior-occipital electrodes—scales with working memory load and tracks individual differences in working memory capacity (Vogel & Machizawa, 2004b). The distractor positivity (P_D) tracks suppression of items from working memory and is also measured using posterior electrodes (Feldmann-Wüstefeld & Vogel, 2018). In light of these and other posterior EEG signals that track working memory processes, the lack of frontal electrode correlations in our models is potentially less surprising.

Tracking brain function

Our results suggest that common and unique signals of cognition are present in temporally precise, but spatially insensitive, EEG signals. This suggests that millisecond-by-millisecond fluctuations in neural processing are unique across individuals and cognitively

meaningful. Previous work has also shown that analogous spatially sensitive signals are present in fMRI functional connectivity networks. These two methods track different neural signatures. But, when considered together, they provide additional evidence that methods that measure statistical dependence between two regions' signal time series, in general, track meaningful variation in brain function.

Despite the robustness of our EEG- and previous fMRI-based fingerprinting results and predictive models, there are certain obstacles that need to be addressed before we can fully accept that these methods track brain function. For example, previous work using fMRI functional connectivity has been criticized because it measures blood oxygenation, which is an indirect measure of neural activity. Due to this, critics have suggested that fMRI functional connectivity fingerprinting could be driven by individual differences in brain structure or vasculature, rather than meaningful differences in brain function (Dubois & Adolphs, 2016; Llera et al., 2019). They additionally suggest that identification and behavioral prediction could be driven by trait-like head motion, which is a challenging confound in fMRI (Nentwich et al., 2020; Siegel et al., 2017; Xifra-Porxas et al., 2020). Our EEG models complement this work by controlling for and ruling out these potential confounds.

EEG—like MRI and all other measures of human brain activity—is also influenced by brain structure. Unlike MRI, however patterns of inter-electrode correlations in the current work measure the topography of electrical brain activity, which is a direct measure of neural activity. EEG is also less influenced by head motion because EEG trials are typically shorter, and trials with interference from head motion are removed from analyses. Therefore, these factors do not influence our ability to identify individuals and predict behavior using inter-electrode correlations. Nevertheless, our EEG analyses come with their own unique challenges. For

example, identification of individuals could be driven by differences in skull thickness across participants. Skull thickness could influence the conduction of neural signals on the scalp, leading to idiosyncratic patterns of electrical activity across people that are unrelated to brain function. We addressed this potential issue of volume conduction by applying a Laplacian transformation, which reduces the adverse effects of volume conduction (Kayser & Tenke, 2015). Even when we account for potential differences in volume conduction on the scalp, we are still able to identify individuals. Overall, our results provide strong evidence that sparse EEG networks can robustly predict cognitive ability. Inter-electrode correlations measured using EEG track meaningful variation in neural activity that is related to individual differences in cognition.

Limitations and future directions

In the current work, we analyze the correlation between two electrodes' trial-evoked EEG signal over time. These analyses do not necessarily provide evidence for communication or causal relationships between, or shared processing in, different electrodes or brain regions. These are separate—but important—questions that should be addressed by future research. One way to begin addressing these questions, which has been utilized in the fMRI literature, would be to characterize inter-electrodes correlations in resting-state (i.e., task-free) EEG data and evaluate their utility for identifying individuals and predicting behavior.

We analyzed EEG data while participants performed variants of a lateralized change detection task. Therefore, it is not clear whether predictions of working memory and general fluid intelligence rely on activity evoked during this specific task state, or whether models would generalize to predict behavior from data collected in other contexts. For example, are behaviorally relevant signals present in intrinsic inter-electrode correlations—that is, EEG

activity observed at rest? Are individual differences in a cognitive process best predicted by inter-electrode correlation observed during a task that taxes that particular process (*c.f.*, Finn et al., 2017)? To address these questions, future research can use EEG data collected during rest and different tasks to predict individual differences in other cognitive abilities, such as attentional control or long-term memory.

One complication with analyzing resting-state EEG data is that our inter-electrode correlation measure may rely on time-locked data that is averaged over many trials. This is in contrast to fMRI functional connectivity analyses, which concatenate trial data. We calculated time-locked averaged EEG data because single-trial data were noisy. Resting-state EEG data are typically collected continuously for a specific amount of time. It does not have specific trials, and, therefore, it is not clear how to overcome low signal-to-noise ratios with continuous resting-state EEG data. Nevertheless, it would be theoretically impactful to determine whether there are behaviorally relevant EEG networks that are present at rest.

Conclusions

We demonstrate that a temporally precise, but sparse network of electrical activity identifies individuals and predicts cognitive abilities. Thus, it is not necessary to measure dense functional networks with relatively high spatial resolution to predict important individual variability in cognition. Instead, cognitively meaningful variability in functional brain organization is also reflected in sparse, high-frequency neural signals. In sum, our analyses, which are temporally sensitive and easy and affordable to implement, provide a new arena in which we can better track and understand important individual differences in cognitive abilities.

GENERAL CONCLUSIONS

There has been a longstanding effort to understand what exactly working memory is and how it is represented in the brain. Most contemporary theories of working memory have embraced the embedded processes model, which proposes that information in working memory is maintained in three distinct layers: the focus of attention, activated long term memory, and long-term memory. Here, I offer a new hypothesis regarding the distinct sub-components of working memory maintenance by providing evidence for dissociable delay period activity. The dissociation between distinct sub-components of working memory further elucidates how information is maintained, with the hope of providing further traction for understanding this central component of intelligent human behavior.

In Chapter 1, I extended the growing evidence that working memory may be implemented via multiple component processes that play distinct functional roles. To do this, I built a computational model of working memory performance that dissociated the contribution of attentional control and working memory storage to the maintenance of information in working memory over time. Previous research has shown that performance on a working memory task declines with the amount of time participants have to maintain information in mind. Previous research has suggested that this decline is attributable to changes in precision as well as sudden loss of item representations. Here, by measuring trial-to-trial variations in performance, I examined an orthogonal distinction between the maximum number of items that an individual can store, and the probability of achieving that maximum (attentional control). Across two experiments, I replicated the finding that performance declines after long retention intervals, as well as past observations that forgetting was due to probabilistic dropping of individual items rather than all-or-none losses of the stored memories. Critically, longer retention intervals did not

reduce the maximum amount of information that could be stored in working memory. Instead, lower attentional control accounted for a decreased probability of maintaining the maximum number of items in working memory. Thus, both memory storage and attentional control contribute to the maintenance of information in working memory. However, longer retention intervals impact working memory storage via fluctuations in attentional control that lower the probability of achieving a stable maximum storage capacity.

In Chapter 2, I tracked neural activity using two time-resolved EEG signals, CDA and alpha power. I found distinct delay-period signatures for an attention task (which required only spatial attention) and a working memory task (which invoked both spatial attention and object storage). Thus, I argued that working memory information is maintained via a collaboration between distinct processes for covert spatial orienting and object-based storage. I would like to highlight that this distinction is not the same as the previously proposed distinction between visual and spatial information (Logie, 1995). Previous research has shown a double dissociation between visual and spatial memory, such that impairments in remembering spatial locations do not always co-occur with impairments in remembering visual appearances of stimuli and vice versa (Darling et al., 2006; Klauer & Zhao, 2004). This distinction between spatial and visual information is based on the featural content of memory, and a large body of research has shown that different features can be behaviorally and neurally dissociated. I propose an orthogonal distinction between signals that index the number of individuated representations (independent of featural content) and signals that track prioritized regions of space (independent of the number of individuated representations). The item-based index (e.g. CDA) is content-free and could be thought of as a pointer-based index that tracks the total number of objects maintained in working memory. The spatial index (e.g. alpha power), on the other hand, tracks the locations of relevant

objects without regard to the number of item-based representations. Both of these indices are independent from the featural content of memory.

In Chapters 3 and 4, I further investigated how spatial representations and item-based storage vary over time. A key challenge for working memory is to maintain relevant information in the face of salient, new information. Therefore, I investigated the impact of irrelevant (Chapters 3 & 4) and relevant (Chapter 4) interruptions on neural representations of space and item-based storage of information already maintained in working memory. I found that brief irrelevant onsets disrupt two distinct online aspects of working memory. Task expectancy and relevance of the interruptions modulated the timing and magnitude of how these two neural signals responded to the interruptions, suggesting that the brain's response to salient, new information is shaped by task context.

In Chapter 5, I demonstrate that a temporally precise, but sparse network of electrical activity identifies individuals and predicts cognitive abilities, including working memory capacity and general fluid intelligence. Previous work in cognitive and network neuroscience has suggested that cognitive abilities, including working memory, emerge from interactions between dozens or hundreds of brain regions in complex, large-scale brain networks (Bressler & Menon, 2010; Rosenberg et al., 2017). Significant research, typically using fMRI, has utilized these networks to describe and predict behavior (Beaty et al., 2018; Finn et al., 2015; Rosenberg et al., 2016; Sripada et al., 2013; Yamashita et al., 2018). Interestingly, we were able to predict these same behaviors using a relatively sparse amount of information. Thus, although cognitive abilities may involve the interaction between large numbers of disparate brain regions, they can also be summarized using a relatively sparse EEG network. Furthermore, this work provides a

new analytic tool in which we can finely track and better understand important individual differences in cognitive abilities.

Understanding how working memories are represented in the brain and why this system is limited could greatly advance our understanding of intelligence and other higher-level cognitive functions. While future work will be necessary to fully substantiate how information in working memory is neurally represented, my work makes a small step forward in this understanding. It reveals the manifold nature of working memory processes and provides a way to predict individual differences in working memory capacity in unseen individuals. The human brain is a powerful cognitive tool that has given rise to exceptional accomplishment, and hopefully one day, it will be able to understand itself.

REFERENCES

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015a). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015b). 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., Robison, M. K., & Vogel, E. K. (2018). Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance Kirsten. *Journal of Cognitive Neuroscience*, *26*(3), 194–198. https://doi.org/10.1162/jocn

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). *Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.06.04.135053

Andrews, A., Ratwani, R., & Trafton, G. (2009). *Recovering From Interruptions: Does Alert Type Matter ?* 409–413.

Ansorge, U., & Heumann, M. (2003). Top-down contingencies in peripheral cuing: The roles of color and location. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 937–948. https://doi.org/10.1037/0096-1523.29.5.937

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

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

Badre, D. (2012). Opening the gate to working memory. *Proceedings of the National Academy of Sciences*, *109*(49), 19878–19879. https://doi.org/10.1073/pnas.1216902109

Balaban, H., & Luria, R. (2016). Integration of Distinct Objects in Visual Working Memory Depends on Strong Objecthood Cues Even for Different-Dimension Conjunctions. *Cerebral Cortex*, *26*(5), 2093–2104. https://doi.org/10.1093/cercor/bhv038

Balaban, H., & Luria, R. (2017a). Neural and Behavioral Evidence for an Online Resetting Process in Visual Working Memory. *The Journal of Neuroscience*, *37*(5), 1225–1239. https://doi.org/10.1523/JNEUROSCI.2789-16.2016

Balaban, H., & Luria, R. (2017b). Neural and Behavioral Evidence for an Online Resetting Process in Visual Working Memory. *The Journal of Neuroscience*, *37*(5), 1225–1239. https://doi.org/10.1523/JNEUROSCI.2789-16.2016

Beatty, J. (1982). *Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources*. *91*(2), 276–292.

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., Fink, A., Qiu, J., Kwapil, T. R., Kane, M. J., & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences*, *115*(5), 1087–1092. https://doi.org/10.1073/pnas.1713532115

Berggren, N., & Eimer, M. (2016). Does Contralateral Delay Activity Reflect Working Memory Storage or the Current Focus of Spatial Attention within Visual Working Memory? *Journal of Cognitive Neuroscience*, *28*(12). https://doi.org/10.1162/jocn

Bisley, J W, Zaksas, D., Droll, J. A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.*, *91*, 286–300.

Bisley, James W, & Goldberg, M. E. (2003). *Neuronal Activity in the Lateral Intraparietal Area and Spatial Attention*. *299*(January).

Bisley, James W., & Goldberg, M. E. (2010). Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*.

Burgess, A. P., & Ali, L. (2002). Functional connectivity of gamma EEG activity is modulated at low frequency during conscious recollection. *International Journal of Psychophysiology*, *46*(2), 91–100. https://doi.org/10.1016/S0167-8760(02)00108-3

Burra, N., & Kerzel, D. (2014). The distractor positivity (Pd) signals lowering of attentional priority: Evidence from event-related potentials and individual differences. *Psychophysiology*, *51*(7), 685–696. https://doi.org/10.1111/psyp.12215

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences*, *108*(27), 1–4. https://doi.org/10.1073/pnas.1104666108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1104666108

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, *14*(11), 506–515. https://doi.org/10.1016/j.tics.2010.09.001

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001

Chatham, C. H., Frank, M. J., & Badre, D. (2014). *Corticostriatal Output Gating during Selection from Working Memory | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.neuron.2014.01.002

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., Golomb, J. D., & Turk-browne, N. B. (2011). *A Taxonomy of External and Internal Attention*. https://doi.org/10.1146/annurev.psych.093008.100427

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

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2010). Mechanisms of working memory disruption by external interference. *Cerebral Cortex*, *20*(4), 859–872. https://doi.org/10.1093/cercor/bhp150

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience Reviews*, *3*.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, and Computers*, *34*(4), 613–617. https://doi.org/10.3758/BF03195489

Cowan, N. (1999). *An embedded-process model of working memory*.

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

Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Scott Saults, J., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, *23*(10), 2852–2863. https://doi.org/10.1162/jocn.2011.21625

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.*, *7*, 415–423.

Damaševičius, R., Maskeliūnas, R., Kazanavičius, E., & Woźniak, M. (2018). Combining Cryptography with EEG Biometrics. *Computational Intelligence and Neuroscience*, *2018*, 1–11. https://doi.org/10.1155/2018/1867548

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, *103*(37), 13848–13853. https://doi.org/10.1073/pnas.0601417103

Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo-spatial working memory. *Journal of Neurology*, *253*(2), 176–180. https://doi.org/10.1007/s00415-005-0944-3

De Fockert, J. W., Rees, G., Frith, C. D., & Lavie1, 3nilli. (2001). The Role of Working Memory in Visual Selective Attention. *Science*, *291*(January), 1803–1806. https://doi.org/10.1126/science.1056496

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*. https://doi.org/10.1038/s41562-019-0606-6

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178. https://doi.org/10.3758/s13423-014-0675-5

Drew, T., & Vogel, E. K. (2008). Neural Measures of Individual Differences in Selecting and Tracking Multiple Moving Objects. *Journal of Neuroscience*, *28*(16), 4183–4191. https://doi.org/10.1523/JNEUROSCI.0556-08.2008

Dubois, J., & Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443. https://doi.org/10.1016/j.tics.2016.03.014

Duncan, J., & Humphreys, G. W. (1989). *Visual Search and Stimulus Similarity*. 26.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483. https://doi.org/10.1016/S0166-2236(00)01633-7

Egeth, H. E., & Yantis, S. (1997). VISUAL ATTENTION: Control, Representation, and Time Course. *Annual Review of Psychology*, *48*(1), 269–297. https://doi.org/10.1146/annurev.psych.48.1.269

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting Visual Attention Between Objects and Locations: Evidence From Normal and Parietal Lesion Subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–177. https://doi.org/10.1037/0096-3445.123.2.161

Engel, S., Glover, G, & Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology: General*, 23.

Feldmann-Wüstefeld, T., Brandhofer, R., & Schubö, A. (2016). Rewarded visual items capture attention only in heterogeneous contexts. *Psychophysiology*, *53*(7), 1063–1073. https://doi.org/10.1111/psyp.12641

Feldmann-Wüstefeld, T., Busch, N. A., & Schubö, A. (2019). Failed Suppression of Salient Stimuli Precedes Behavioral Errors. *Journal of Cognitive Neuroscience*, *32*(2), 367–377. https://doi.org/10.1162/jocn_a_01502

Feldmann-Wüstefeld, T., Busch, N. A., & Schubö, A. (2020). Failed Suppression of Salient Stimuli Precedes Behavioral Errors. *Journal of Cognitive Neuroscience*, *32*(2), 367–377. https://doi.org/10.1162/jocn_a_01502

Feldmann-Wüstefeld, T., & Schubö, A. (2013). Context homogeneity facilitates both distractor inhibition and target enhancement. *Journal of Vision*, *13*(2013), 1–12. https://doi.org/10.1167/13.3.11.doi

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497. https://doi.org/10.1111/psyp.12514

Feldmann-Wüstefeld, T., & Vogel, E. K. (2018). Neural Evidence for the Contribution of Active Suppression During Working Memory Filtering. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhx336

Feldmann-Wüstefeld, T., Vogel, E. K., & Awh, E. (2018). Contralateral Delay Activity Indexes Working Memory Storage, Not the Current Focus of Spatial Attention Tobias Feldmann-

Wüstefeld, Edward K. Vogel, and Edward Awh Abstract. *Journal of Cognitive Neuroscience2*, 1–11. https://doi.org/10.1162/jocn

Fellrath, J., Mottaz, A., Schnider, A., Guggisberg, A. G., & Ptak, R. (2016). Theta-band functional connectivity in the dorsal fronto-parietal network predicts goal-directed attention. *Neuropsychologia*, *92*, 20–30. https://doi.org/10.1016/j.neuropsychologia.2016.07.012

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & 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

Folk, C. L., Leber, A. B., & Egeth, H. E. (2008). Top-down control settings and the attentional blink: Evidence for nonspatial contingent capture. *Visual Cognition*, *16*(5), 616–642. https://doi.org/10.1080/13506280601134018

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 847–858. https://doi.org/10.1037/0096-1523.24.3.847

Folk, C. L., & Remington, R. W. (2015). Unexpected abrupt onsets can override a top-down set for color. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1153–1165. https://doi.org/10.1037/xhp0000084

Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017a). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, *27*(20), 3216-3223.e6. https://doi.org/10.1016/j.cub.2017.09.031

Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017b). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, *27*(20), 3216-3223.e6. https://doi.org/10.1016/j.cub.2017.09.031

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

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2015). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, *115*(1), 168–177. https://doi.org/10.1152/jn.00860.2015

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2017). Alpha-Band Oscillations Enable Spatially and Temporally Resolved Tracking of Covert Spatial Attention. *Psychological Science*, *28*(7), 929–941. https://doi.org/10.1177/0956797617699167

Fougnie, D. (2008). The relationship between attention and working memory. In *New Research on Short-Term Memory*. https://doi.org/10.3389/conf.fnhum.2011.207.00576

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Fukuda, K., Kang, M.-S., & Woodman, G. F. (2016). Distinct neural mechanisms for spatially lateralized and spatially global visual working memory representations. *Journal of Neurophysiology*, *116*(4), 1715–1727. https://doi.org/10.1152/jn.00991.2015

Fukuda, K., Mance, I., & Vogel, E. K. (2015). Alpha Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory. *Journal of Neuroscience*, *35*(41), 14009–14016. https://doi.org/10.1523/JNEUROSCI.5003-14.2015

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science*, *22*(3), 361–368. https://doi.org/10.1177/0956797611398493

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

Funahashi, S., Chafee, M. V., & Goldman-Rakic, P. S. (1993). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*, *365*(6448), 753–756. https://doi.org/10.1038/365753a0

Gaspar, J. M., & McDonald, J. J. (2014a). Suppression of Salient Objects Prevents Distraction in Visual Search. *Journal of Neuroscience*, *34*(16), 5658–5666. https://doi.org/10.1523/JNEUROSCI.4161-13.2014

Gaspar, J. M., & McDonald, J. J. (2014b). Suppression of Salient Objects Prevents Distraction in Visual Search. *Journal of Neuroscience*, *34*(16), 5658–5666. https://doi.org/10.1523/JNEUROSCI.4161-13.2014

Gaspar, John M, & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. *The Journal of Neuroscience*, *34*(16), 5658–5666. https://doi.org/10.1523/JNEUROSCI.4161-13.2014

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015a). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science*, *26*(11), 1740–1750. https://doi.org/10.1177/0956797615597913

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015b). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science*, *26*(11), 1740–1750. https://doi.org/10.1177/0956797615597913

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2016). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception, & Psychophysics*, 1–18. https://doi.org/10.3758/s13414-016-1209-1

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2017). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception, & Psychophysics*, *79*(1), 45–62. https://doi.org/10.3758/s13414-016-1209-1

Gaspelin, N., & Luck, S. J. (2018). Combined Electrophysiological and Behavioral Evidence for the Suppression of Salient Distractors. *Journal of Cognitive Neuroscience*, *30*(9), 1265–1280. https://doi.org/10.1162/jocn_a_01279

Gratton, C., Lauman, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gillmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*. https://doi.org/10.1016/j.neuron.2018.03.035

Greene, D. J., Koller, J. M., Hampton, J. M., Wesevich, V., Van, A. N., Nguyen, A. L., Hoyt, C. R., McIntyre, L., Earl, E. A., Klein, R. L., Shimony, J. S., Petersen, S. E., Schlaggar, B. L., Fair, D. A., & Dosenbach, N. U. F. (2018). Behavioral interventions for reducing head motion during MRI scans in children. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.01.023

Gunseli, E., Fahrenfort, J. J., van Moorselaar, D., Daoultzis, K. C., Meeter, M., & Olivers, C. N. L. (2019). EEG dynamics reveal a dissociation between storage and selective attention within working memory. *Scientific Reports*.

Hagemann, D., Hewig, J., Walter, C., & Naumann, E. (2008). Skull thickness and magnitude of EEG alpha activity. *Clinical Neurophysiology*, *119*(6), 1271–1280. https://doi.org/10.1016/j.clinph.2008.02.010

Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E., & Vogel, E. K. (2018). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *BioRxiv*.

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., Edward, A., & Edward K, V. (2020). *Working Memory: The state of the science*. Oxford University Press.

Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2019). Perturbing Neural Representations of Working Memory with Task-irrelevant Interruption. *Journal of Cognitive Neuroscience*, *32*(3), 558–569. https://doi.org/10.1162/jocn_a_01481

Harrison, S. A., & Tong, F. (2009). *Visual Areas*. *458*(7238), 632–635. https://doi.org/10.1038/nature07832.Decoding

Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*, *21*(4), 760–775. https://doi.org/10.1162/jocn.2009.21039

Hollingworth, A., & Beck, V. M. (2016). Memory-Based Attention Capture when Multiple Items Are Maintained in Visual Working Memory. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(7), 911–917. https://doi.org/10.1037/xhp0000230

Horstmann, G. (2005). Attentional capture by an unannounced color singleton depends on expectation discrepancy. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 1039–1060. https://doi.org/10.1037/0096-1523.31.5.1039

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*(4), 1963–1968. https://doi.org/10.1152/jn.00978.2009

Jonides, J., Lacey, S. C., & Nee, D. E. (2015). Processes of Working Memory in Mind and Brain. *Society*, *14*(1), 12–15. https://doi.org/10.1111/j.0963-7214.2005.00323.x

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

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

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). *The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception*. 10.

Karamzadeh, N., Medvedev, A., Azari, A., Gandjbakhche, A., & Najafizadeh, L. (2013). Capturing dynamic patterns of task-based functional connectivity with EEG. *NeuroImage*, *66*, 311–317. https://doi.org/10.1016/j.neuroimage.2012.10.032

Kayser, J., & Tenke, C. E. (2015). On the benefits of using surface Laplacian (Current Source Density) methodology in electrophysiology. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *97*(3), 171–173. https://doi.org/10.1016/j.ijpsycho.2015.06.001

Kessler, D., Angstadt, M., & Sripada, C. (2016). Growth Charting of Brain Connectivity Networks and the Identification of Attention Impairment in Youth. *JAMA Psychiatry*, *73*(5), 481. https://doi.org/10.1001/jamapsychiatry.2016.0088

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, *133*(3), 355–381. https://doi.org/10.1037/0096-3445.133.3.355

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Lamprecht, R., & LeDoux, J. (2004). Structural plasticity and memory. *Nature Reviews Neuroscience*, *5*(1), 45–54. https://doi.org/10.1038/nrn1301

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

Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *NeuroImage*, *156*, 166–173. https://doi.org/10.1016/j.neuroimage.2017.05.016

Llera, A., Wolfers, T., Mulders, P., & Beckmann, C. F. (2019). Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *ELife*, *8*, e44443. https://doi.org/10.7554/eLife.44443

Logie, R. H. (1995). Visuo-spatial working memory. In *Exploring Working Memory*. https://doi.org/10.4324/9781315111261-17

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

Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 902–912. https://doi.org/10.1037/0096-1523.28.4.902

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016a). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience and Biobehavioral Reviews*, *34*(5), 352–359. https://doi.org/10.1177/0963721414541462.Self-Control

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016b). 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

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. https://doi.org/10.1016/j.neuropsychologia.2010.11.031

Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. *Cerebral Cortex*, *26*(1), 166–179. https://doi.org/10.1093/cercor/bhu189

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention: Behavioral decoding of attended memory items. *Annals of the New York Academy of Sciences*, *1424*(1), 256–267. https://doi.org/10.1111/nyas.13647

Marini, F., Chelazzi, L., & Maravita, A. (2013). The costly filtering of potential distraction: Evidence for a supramodal mechanism. *Journal of Experimental Psychology: General*, *142*(3), 906–922. https://doi.org/10.1037/a0029905

Maxwell, J. W., Gaspelin, N., & Ruthruff, E. (2020). No identifcation of abrupt onsets that capture attention: Evidence against a unifed model of spatial attention. *Psychological Research*.

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*(3), 531–536. https://doi.org/10.1007/s00221-007-1002-4

Mcdonald, J. J., Green, J. J., Jannati, A., & Di Lollo, V. (2013). On the electrophysiological evidence for the capture of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 849–860. https://doi.org/10.1037/a0030510

McVay, J. C., & Kane, M. J. (2009). Conducting the Train of Thought: Working Memory Capacity, Goal Neglect, and Mind Wandering in an Executive-Control Task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10.1037/a0014104

McVay, J. C., & Kane, M. J. (2010). Does Mind Wandering Reflect Executive Function or Executive Failure? Comment on and. *Psychological Bulletin*, *136*(2), 188–207. https://doi.org/10.1037/a0018298

Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience*, *27*(8), 1471–1491. https://doi.org/10.1162/jocn_a_00810

Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., & Fair, D. A. (2014). Connectotyping: Model Based Fingerprinting of the Functional Connectome. *PLoS ONE*, *9*(11), e111048. https://doi.org/10.1371/journal.pone.0111048

Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional Capture by Salient Color Singleton Distractors Is Modulated by Top-Down Dimensional Set. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 1–16. https://doi.org/10.1037/0096-1523.35.1.1

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention Restores Discrete Items to Visual Short-Term Memory. *Psychological Science*, *24*(4), 550–556. https://doi.org/10.1177/0956797612457782

Nentwich, M., Ai, L., Madsen, J., Telesford, Q. K., Haufe, S., Milham, M. P., & Parra, L. C. (2020). Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2020.117001

Oberauer, K. (2002). Access to Information in Working Memory: Exploring the Focus of Attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(3), 411–421. https://doi.org/10.1037//0278-7393.28.3.411

O'Halloran, L., Cao, Z., Ruddy, K., Jollans, L., Albaugh, M. D., Aleni, A., Potter, A. S., Vahey, N., Banaschewski, T., Hohmann, S., Bokde, A. L. W., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivières, S., Flor, H., Frouin, V., Gowland, P., Heinz, 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

Olivers, C. N. L. (2008). Interactions between visual working memory and visual attention. *Frontiers in Biological Science*, 1182–1191.

Olivers, C. N. L., & Eimer, M. (2011). On the difference between working memory and attentional set. *Neuropsychologia*, *49*(6), 1553–1558. https://doi.org/10.1016/j.neuropsychologia.2010.11.033

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*(1–9). https://doi.org/10.1155/2011/156869

O'Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the*

National Academy of Sciences, *107*(16), 7580–7585. https://doi.org/10.1073/pnas.0913113107

Park, H.-J., & Friston, K. (2013). Structural and Functional Brain Networks: From Connections to Cognition. *Science*, *342*(6158), 1238411–1238411. https://doi.org/10.1126/science.1238411

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry*.

Poole, V. N., Robinson, M. E., Singleton, O., DeGutis, J., Milberg, W. P., McGlinchey, R. E., Salat, D. H., & Esterman, M. (2016). Intrinsic functional connectivity predicts individual differences in distractibility. *Neuropsychologia*, *86*, 176–182. https://doi.org/10.1016/j.neuropsychologia.2016.04.023

Postle, B. R. (2006). *Working Memort as an Emergent Property of the Mind and Brain*. *100*(2), 130–134. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory and Cognition*, *33*(2), 203–212. https://doi.org/10.3758/BF03195309

Prinzmetal, W., Zvinyatskovskiy, A., Gutierrez, P., & Dilem, L. (2009). Voluntary and involuntary attention have different consequences: The effect of perceptual difficulty. *Quarterly Journal of Experimental Psychology*, *62*(2), 352–369. https://doi.org/10.1080/17470210801954892

Rademaker, R. L., Park, Y. E., Sack, A. T., & Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*.

Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *European Journal of Neuroscience*, *25*(2), 603–610. https://doi.org/10.1111/j.1460-9568.2007.05278.x

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Mayering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing attention with predictive network models. *Trends in Cognitive Sciences*, *21*(4), 290–302. https://doi.org/10.1016/j.tics.2017.01.011

Rosenberg, M. D., 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

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., Glennon, M., Gerloff, C., Birbaumer, N., & Hummel, F. C. (2009). Brain Oscillatory Substrates of Visual Short-Term Memory Capacity. *Current Biology*, *19*(21), 1846–1852. https://doi.org/10.1016/j.cub.2009.08.062

Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A Common Neural Mechanism for Preventing and Terminating the Allocation of Attention. *Journal of Neuroscience*, *32*(31), 10725–10736. https://doi.org/10.1523/JNEUROSCI.1864-12.2012

Sawaki, Risa, & Luck, S. J. (2012). Active suppression of distractors that match the contents of visual working memory Risa. *Visual Cognition*, *19*(7), 1–14. https://doi.org/10.1080/13506285.2011.603709.Active

Schneegans, S., & Bays, P. M. (2018). Drift in neural population activity causes working memory to deteriorate over time. *The Journal of Neuroscience*, 3440–17. https://doi.org/10.1523/JNEUROSCI.3440-17.2018

Schubert, A. L., Hagemann, D., & Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology: General*.

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

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*, *38*(6), 1716–1717. https://doi.org/10.1523/JNEUROSCI.1716-17.2017

Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., & Snyder, A. Z. (2017). Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cerebral Cortex*, *27*(9), 4492–4502. https://doi.org/10.1093/cercor/bhw253

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348. https://doi.org/10.1016/j.tics.2008.05.007

Sripada, C. S., Phan, K. L., Labuschagne, I., Welsh, R., Nathan, P. J., & Wood, A. G. (2013). Oxytocin enhances resting-state connectivity between amygdala and medial frontal cortex. *International Journal of Neuropsychopharmacology*, *16*(2), 255–260. https://doi.org/10.1017/S1461145712000533

Steinhauer, S., & Hakerem, G. (1992). The Pupillary Response in Cognitive Psychophysiology and Schizophrenia. *Annals of the New York Academy of Sciences*, *658*(1), 182–204. https://doi.org/10.1111/j.1749-6632.1992.tb22845.x

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

Tas, A. C., Luck, S. J., & Hollingworth, A. (2016). The relationship between visual attention and visual working memory encoding: A dissociation between covert and overt orienting. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1121–1138. https://doi.org/10.1037/xhp0000212

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. *Journal of Neuroscience*, *26*(37), 9494–9502. https://doi.org/10.1523/JNEUROSCI.0875-06.2006

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754.

Töllner, T., Müller, H. J., & Zehetleitner, M. (2012). Top-down dimensional weight set determines the capture of visual attention: Evidence from the PCN component. *Cerebral Cortex*, *22*(7), 1554–1563. https://doi.org/10.1093/cercor/bhr231

Tsubomi, H., Fukuda, K., Watanabe, K., & Vogel, E. K. (2013). Neural Limits to Representing Objects Still within View. *Journal of Neuroscience*, *33*(19), 8257–8263. https://doi.org/10.1523/JNEUROSCI.5348-12.2013

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working Memory Delay Activity Predicts Individual Differences in Cognitive Abilities. *Journal of Cognitive Neuroscience*, *27*(5), 853–865. https://doi.org/10.1162/jocn_a_00765

van Dijk, H., van der Werf, J., Mazaheri, A., Medendorp, W. P., & Jensen, O. (2010). Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. *Proceedings of the National Academy of Sciences*, *107*(2), 900–905. https://doi.org/10.1073/pnas.0908821107

van Moorselaar, D., Foster, J. J., Sutterer, D. W., Theeuwes, J., Olivers, C. N. L., & Awh, E. (2017a). Spatially Selective Alpha Oscillations Reveal Moment-by-Moment Trade-offs between Working Memory and Attention. *Journal of Cognitive Neuroscience*, *30*(2), 256–266. https://doi.org/10.1162/jocn

van Moorselaar, D., Foster, J. J., Sutterer, D. W., Theeuwes, J., Olivers, C. N. L., & Awh, E. (2017b). Spatially Selective Alpha Oscillations Reveal Moment-by-Moment Trade-offs

between Working Memory and Attention. *Journal of Cognitive Neuroscience*, *30*(2), 256–266. https://doi.org/10.1162/jocn_a_01198

van Moorselaar, D., Foster, J. J., Sutterer, D. W., Theeuwes, J., Olivers, C. N. L., & Awh, E. (2018). Spatially Selective Alpha Oscillations Reveal Moment-by-Moment Trade-offs between Working Memory and Attention. *Journal of Cognitive Neuroscience*, *26*(3), 194–198. https://doi.org/10.1162/jocn

van Moorselaar, D., Gunseli, E., Theeuwes, J., & N. L. Olivers, C. (2015). The time course of protecting a visual memory representation from perceptual interference. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.01053

Vogel, E. K., & Machizawa, M. G. (2004a). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Vogel, E. K., & Machizawa, M. G. (2004b). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451. https://doi.org/10.1037/0096-1523.32.6.1436

Weaver, M. D., Hickey, C., & Zoest, W. van. (2017). The impact of salience and visual working memory on the monitoring and control of saccadic behavior: An eye-tracking and EEG study. *Psychophysiology*, *54*(4), 544–554. https://doi.org/10.1111/psyp.12817

Williams, M., & Woodman, G. F. (2012). Directed Forgetting and Directed Remembering in Visual Working Memory. *Journal Of Experimental Psychology-Learning Memory And Cognition*. https://doi.org/10.1037/a0027389

Williams, M., & Woodman, G. F. (2013). *Directed Forgetting and Directed Remembering in Visual Working Memory*. *38*(5). https://doi.org/10.1037/a0027389.Directed

Wolff, M. J., Jochim, J., Akyurek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working memory guided behaviour. *Nature Neurosci*., 1–35.

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Woodman, G. F., Arita, J. T., & Luck, S. J. (2009). *A Cuing Study of the N2pc Component: An Index of Attentional Deployment to Objects Rather Than Spatial Locations*. *1297*, 101–111. https://doi.org/10.1016/j.brainres.2009.08.011.A

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, *14*(4–8), 808–830. https://doi.org/10.1080/13506280500197397

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *The Journal of Neuroscience*, *20*, 1–6. https://doi.org/Rc63

Xifra-Porxas, A., Kassinopoulos, M., & Mitsis, G. D. (2020). *Physiological and head motion signatures in static and time-varying functional connectivity and their subject discriminability* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.02.04.934554

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.

Yamashita, M., Yoshihara, Y., Hashimoto, R., Yahata, N., Ichikawa, N., Sakai, Y., Yamada, T., Matsukawa, N., Okada, G., Tanaka, S. C., Kasai, K., Kato, N., Okamoto, Y., Seymour, B., Takahashi, H., Kawato, M., & Imamizu, H. (2018). A prediction model of working memory across health and psychiatric disease using whole-brain functional connectivity. *ELife*, *7*, e38844. https://doi.org/10.7554/eLife.38844

Yantis, S., & Jonides, J. (1984). Abrupt Visual Onsets and Selective Attention: Evidence From Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 21.

Zhang, W., & Luck, S. J. (2009a). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

Zhang, W., & Luck, S. J. (2009b). Sudden Death and Gradual Decay in Visual Working Memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

Zivony, A., & Lamy, D. (2018). Contingent attentional engagement: Stimulus- and goal-driven capture have qualitatively different consequences. *Psychological Science*.