



HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Neuroimage. 2018 May 01; 171: 148–164. doi:10.1016/j.neuroimage.2018.01.002.

Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension

Francesca C. Fortenbaugh^{1,2,3,4}, David Rothlein^{1,2,3,5}, Regina McGlinchey^{1,3,4}, Joseph DeGutis^{1,2,3,6}, and Michael Esterman^{1,2,3,5}

¹Neuroimaging Research for Veterans (NeRVe) Center, VA Boston Healthcare System, Boston, MA 02130

²Boston Attention & Learning Laboratory, VA Boston Healthcare System, Boston, MA 02130

³Geriatric Research, Education, & Clinical Center (GRECC), VA Boston Healthcare System, Boston, MA 02130

⁴Department of Psychiatry, Harvard Medical School, Boston, MA 02115

⁵Department of Psychiatry, Boston University School of Medicine, Boston, MA 02118

⁶Department of Medicine, Harvard Medical School, Boston, MA 02115

Abstract

Novel paradigms have allowed for more precise measurements of sustained attention ability and fluctuations in sustained attention over time, as well as the neural basis of fluctuations and lapses in performance. However, in recent years, concerns have arisen over the replicability of neuroimaging studies and psychology more broadly, particularly given the typically small sample sizes. One recently developed paradigm, the gradual-onset continuous performance task (gradCPT) has been validated behaviorally in large samples of participants. Yet neuroimaging studies investigating the neural basis of performance on this task have only been collected in small samples. The present study completed both a robust replication of the original neuroimaging findings and extended previous results from the gradCPT task using a large sample of 140 Veteran participants. Results replicate findings that fluctuations in attentional stability are tracked over time by BOLD activity in task positive (e.g., dorsal and ventral attention networks) and task negative (e.g., default network) regions. Extending prior results, we relate this coupling between attentional stability and on-going brain activity to overall sustained attention ability and demonstrate that this coupling strength, along with across-network coupling, could be used to predict individual differences in performance. Additionally, the results extend previous findings by demonstrating that temporal dynamics across the default and dorsal attention networks are associated with lapse-likelihood on subsequent trials. This study demonstrates the reliability of the

Corresponding Author: Francesca Fortenbaugh, Ph.D., 150 S. Huntington Avenue, Boston, MA 02130, Francesca_Fortenbaugh@hms.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

gradCPT, and underscores the utility of this paradigm in understanding attentional fluctuations, as well as individual variation and deficits in sustained attention.

Keywords

continuous performance task; vigilance; fMRI; default mode network; dorsal attention network

1. Introduction

Over the course of a day, individuals consistently employ and sustain attention to a multitude of tasks. Whether driving to work or reading a paper, the ability to maintain focused voluntary attention on a single task is a critical cognitive function that allows individuals to effectively interact with their environments and complete goals. Given that the ability to sustain attention can profoundly impact many other cognitive and sensory functions (Barkley, 1997; Fortenbaugh, Robertson, & Esterman, 2017; Sarter, Givens, & Bruno, 2001; H. Silver & Feldman, 2005), characterizing sustained attention abilities has been an active area of research for decades, with some studies focused on understanding fluctuations of or decrements in sustained attention ability across time in healthy observers (Berardi, Parasuraman, & Haxby, 2001; Carriere, Cheyne, Solman, & Smilek, 2010; Esterman, Rosenberg, & Noonan, 2014; Fortenbaugh et al., 2015; Levy, 1980; Mackworth, 1948; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997; Rosenberg et al., 2016; Sarter et al., 2001; Staub, Doignon-Camus, Bacon, & Bonnefond, 2014; Staub, Doignon-Camus, Després, & Bonnefond, 2013), and others focused on characterizing deficits in sustained attention ability associated with psychiatric and neurological disorders (Altpeter, Mackeben, & Trauzettel-Klosinski, 2000; Barkley, 1997; Clark, Iversen, & Goodwin, 2002; Forster, Nunez Elizalde, Castle, & Bishop, 2015; Park, Hood, Shah, Fogg, & Wyatt, 2012; Rosenberg et al., 2016; Van Vleet & DeGutis, 2013).

In recent years, researchers have made substantial progress in characterizing the neural networks involved in sustained attention (Clayton, Yeung, & Cohen Kadosh, 2015; Esterman, Noonan, Rosenberg, & DeGutis, 2013; Esterman, Rosenberg, et al., 2014; Fortenbaugh, DeGutis, & Esterman, 2017; Langner & Eickoff, 2013; Lawrence, Ross, Hoffmann, Garavan, & Stein, 2003; Rosenberg et al., 2016; Sarter et al., 2001). Enabling this progress has been the development of novel tasks and analytic methods that allow for more precise measurements of sustained attention ability and induce more failures in sustained attention over shorter testing periods, increasing sensitivity to individual differences, as well as behavioral relationships with brain activity/connectivity. One commonly used paradigm in this literature is the not-X continuous performance task, requiring participants to frequently respond to non-target stimuli and infrequently withhold responses to rare target stimuli. This task allows measurements of sustained attention and vigilance decrements to be obtained over much shorter periods of time than other tasks, which involve responses only to infrequent target events, while at the same time sampling behavior at a high rate. These include the commonly used Sustained Attention to Response Task (SART) (Robertson et al., 1997), and the Gradual Onset Continuous Performance Task (gradCPT) (Esterman et al., 2013), as well as many other innovative variations (Helton &

Russell, 2011; Kucyi, Hove, Esterman, Hutchison, & Valera, 2017; Shalev, Ben-Simon, Mevorach, Cohen, & Tsal, 2011; Temple et al., 2000). One unique feature that was introduced in the gradCPT is the use of gradual transitions from one trial image to the next, eliminating the abrupt offsets and onsets of stimuli between trials that can serve to orient involuntary attention toward the display (Fortenbaugh et al., 2015; Rosenberg, Noonan, DeGutis, & Esterman, 2013). The removal of these abrupt offsets/onsets makes the task more dependent on endogenous attentional control both behaviorally and with regard to fluctuations in the fMRI signal.

Previous studies using the gradCPT have leveraged its sensitive and data-rich behavioral output to identify and examine a number of behavioral and neural indicators of both instantaneous attentional state and overall sustained attention ability. For example, results from the original gradCPT study (Esterman et al., 2013) showed that while the default, dorsal attention, and sensory regions demonstrated characteristic task-negative and task-positive BOLD responses to the onset of target (mountain) scenes, preparatory (pre-trial) activity in these regions was also associated with subsequent accuracy. Specifically, greater activity in stimulus-selective parahippocampal place area (PPA) and dorsal attention network (DAN) was associated with subsequent accuracy, while greater activity in the default mode network (DMN) was associated with subsequent errors. These results are consistent with other studies that indicate that ongoing DMN activity may reflect task-unrelated thoughts such as mind wandering (Andrews - Hanna, Smallwood, & Spreng, 2014; Broyd et al., 2009; Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016; Greicius, Krasnow, Reiss, & Menon, 2003; Mason et al., 2007; Raichle et al., 2001), and that ongoing sensory/DAN activation may reflect ongoing attention to task-related stimuli (Corbetta & Shulman, 2002; Posner & Peterson, 1990; M. A. Silver & Kastner, 2009). In addition to examining activity surrounding rare target events, the original study by Esterman et al. (2013) computed a continuous dynamic metric of reaction time variability, which revealed that sustained performance can be characterized by at least two states: when participants are “in the zone” versus “out of the zone”. Periods of being “in the zone” are defined based on low reaction time variability to frequent non-target stimuli (e.g., images of city scenes) while “out of the zone” is defined as periods of higher reaction time variability. Analyses of in-the-zone versus out-of-the-zone periods revealed that accuracy was higher (fewer errors of commission and omission) during in-the-zone periods. In contrast to preparatory activation associated with target accuracy, fluctuations between these attentional states were coupled with on-going brain activity in the default mode network (DMN) such that greater activation was associated with being in the zone. The dorsal attention network (DAN) exhibited the opposite relationship--greater activity when out of the zone. Subsequent studies corroborated and extended the findings about these relationships, indicating greater task-negative activation when in the zone and greater task-positive activation when out of the zone (in dorsal and ventral attention regions; Kucyi 2016; Esterman 2016; Esterman 2014). Further, these patterns interacted with preparatory activity before targets (correct vs. incorrect) such that task-positive effects were stronger out of the zone and task-negative effects were stronger in the zone. This led to the hypothesis that optimal attentional states are not simply reflected by task-positive and task-negative activation alone. Specifically, attentional fluctuations can be described with multiple behavioral markers- accuracy, mind wandering,

RT variability, and motivational state- each of which may have independent and even opposing contributions to brain activity across large-scale brain networks. This dichotomous relationship between the neural markers of accuracy and variability suggest that, for tasks that require constant engagement across extended periods of time, prolonged suppression of DMN and/or activation of DAN may not be sustainable and may undermine attentional stability over time. Thus, in relation to the observed variability coupling with brain activity, moderate increases in DMN activity during “in the zone” periods and decreases in task-positive attentional control regions such as the DAN may indicate a more distributed and/or efficient attentional state that can be maintained over periods of time. One unanswered question regarding this somewhat surprising variability-brain coupling is whether the degree to which DMN and DAN are coupled with fluctuations in variability is related to overall attention ability across participants. Specifically, do participants with better sustained performance show greater coupling, supporting the idea that this coupling helps maintain a balance or optimal activation across task-negative and task-positive networks.

Since the initial publication, the gradCPT and its variants have been used to further characterize sustained attention both in neurotypical (Esterman et al., 2016; Esterman et al., 2015; Esterman, Poole, Liu, & DeGutis, 2017; Esterman, Reagan, Liu, Turner, & DeGutis, 2014; Esterman, Rosenberg, et al., 2014; Kucyi, Esterman, Riley, & Valera, 2016; Kucyi et al., 2017; Rosenberg et al., 2013) and clinical populations (Auerbach et al., 2014; Fortenbaugh, Corbo, et al., 2017; Rosenberg et al., 2016). Further, this task has been used to explore variation in sustained attention associated with age, gender, sociocultural factors, and time of day (Fortenbaugh et al., 2015; Riley, Esterman, Fortenbaugh, & DeGutis, in press; Riley et al., 2016). Performance, as well as the neural correlates of fluctuations in accuracy (preparatory activity) and variability (in/out of the zone), have been shown to be modulated by motivation and reward (Esterman et al., 2016; Esterman et al., 2017; Esterman, Reagan, et al., 2014). In clinical samples, behavioral performance on the gradCPT has been associated with PTSD, depression, and early life trauma (Auerbach et al., 2014; DeGutis et al., 2015; Fortenbaugh, Corbo, et al., 2017). Analyses of functional connectivity during the task, although outside the scope of this paper, are sensitive to individual differences in performance, early life trauma, and ADHD (Fortenbaugh, Corbo, et al., 2017; Rosenberg et al., 2016).

There were two goals of the present study. First, we sought to replicate the core original findings from the Esterman et al. (2013) study. While multiple studies have utilized the gradCPT paradigm to ask novel questions, to date, the core findings regarding the relationship between ongoing activity in the DAN and DMN to ongoing attentional stability and pretrial activity in these regions to attentional lapses, has not been replicated. This is important as questions have arisen in recent years regarding the extent to which many findings in psychology and neuroscience replicate and generalize to larger samples that are not limited to self-selecting college students, have a greater range in baseline intelligence/ cognitive functioning, and are more representative of the general population as a whole (Boekel et al., 2015; Button et al., 2013; Open Science Collaboration, 2015; Poldrack et al., 2017). Within the neurosciences, one of the primary issues that has been raised regarding findings from functional magnetic resonance imaging (fMRI) studies, is the low power that is associated in part with small sample sizes (Button et al., 2013). Given the diverse set of

inferences being drawn from gradCPT, it is critical to determine whether the core behavioral and neural findings are both replicable and robust to changes in sampling population. The behavioral aspects of the gradCPT, including overall performance, the relationship between variables, as well as the reliability of each variable, have been validated in a large, heterogeneous sample of participants (>10,000). In terms of the fMRI findings, variability-BOLD correlations have been replicated in several gradCPT studies (Esterman et al., 2017; Kucyi et al., 2017), as well as in other sustained attention tasks (Johnson et al., 2015; Kucyi et al., 2017; Rosenberg, Finn, Constable, & Chun, 2015). However, these studies have only been assessed in small groups of healthy, young participants of a relatively restricted range of ages, demographics, and health status (Esterman et al., 2013; Esterman, Rosenberg, et al., 2014). In order to address these issues, and to determine the extent to which the original gradCPT findings regarding brain-behavior relationships replicate to a larger sample, with more variability across a range of demographic factors, the present study assessed performance on the gradCPT with concurrent functional magnetic resonance imaging (fMRI) in a large sample of 140 Veterans.

The second goal of the current study was to extend previous findings from the original Esterman et al. (2013) study in three important ways. First, we used whole-brain voxel level analyses rather than the region-of-interest approach originally used to more fully characterize the evoked responses of all three response types possible on gradCPT task, as well as contrasts between different transient events. This revealed overlapping (e.g., salience network) and distinct (e.g., ventral visual cortex) activation markers of attentional lapses. Further, we have extensively expanded the lapses precursor analysis, by exploring the whole brain, as well as a longer trajectory of activation preceding errors. We find that areas outside of our a priori sensory, DAN, and DMN regions predict errors, and also show that error trajectories begin up to 12.8 seconds before a lapse of attention. In addition to these group-level expanded analyses, we further explore the variance time course (in/out of zone BOLD coupling) in several ways that help elucidate the role of this variability/brain coupling in sustained attention. While it has been previously shown that BOLD activity in several regions tracks reaction time stability as measured with the variance time course (VTC), the degree to which the strength of this coupling affects overall performance or is predictive of individual differences in performance has not been tested. Leveraging the large sample size of the current study, we demonstrate here across two analyses that VTC coupling strength is associated with and can be used to predict overall performance on the task.

2. Methods

2.1 Participants

The initial participant sample included 157 Veterans from Operation Enduring Freedom/Operation Iraqi Freedom/Operation New Dawn (OEF/OIF/OND) who were recruited from the Translational Research Center for Traumatic Brain Injury and Stress Disorders (TRACTS) at the Veterans Affairs Boston Healthcare System. Details regarding the larger TRACTS cohort has been described in detail elsewhere (Lippa et al., 2015; McGlinchey, Milberg, Fonda, & Fortier, 2017). Exclusion criteria for TRACTS includes: (a) history of neurological illness (other than traumatic brain injury (TBI)); (b) history of seizures; (c)

current diagnosis of schizophrenia spectrum or other psychotic disorders (not related to PTSD); (d) current active suicidal and/or homicidal ideation, intent, or plan requiring crisis intervention; or (e) cognitive disorder due to general medical condition other than TBI. Participants in the present sample completed the current task during MRI scanning as part of a larger battery of tests that includes clinical interviews and neuropsychological testing. The TRACTS MRI protocol, completed at the end of the 8–10 hour testing day, includes two structural MRI scans, a diffusion tensor imaging (DTI) scan, and two 6-min resting state scans. The gradCPT functional scan was appended to the end of the MRI protocol between January 2013 and March 2015. Due to limited scanning time, only one run of the gradCPT was collected in contrast to the original study which collected multiple runs (Esterman et al., 2013). The Institutional Review Board of Human Studies Research at the VA Boston Healthcare System approved all research procedures. All participants provided informed consent and were reimbursed for their time and travel expenses.

For the present study, veterans with a history of moderate or severe TBI were excluded from the final sample ($N=7$), as moderate and severe TBI has been consistently shown to impair cognition (Robertson et al., 1997; Schretlen & Shapiro, 2003; Slovarp, Azuma, & LaPointe, 2012). Data from an additional participant was excluded due to technical difficulties (no responses recorded during task). An additional 5 participants were excluded due to excessive motion during the functional gradCPT scan run. This included 2 participants who moved over 5mm over the course of the 8-min run, and 3 participants who had more than 30% of their data points censored using the criteria outlined below. Finally, following the methodology of Fortenbaugh et al. (2015), we assessed whether participants showed significant periods of inactivity, defined as 30 second periods or greater with no response to the task (“tune outs”). Lack of responses over such large periods of time (i.e. >37 consecutive trials) could be due to lack of engagement on the task by participants. A total of 4 participants were excluded using this criterion leading to a final sample size of 140 participants (131 males; 32.1 ± 7.7 years of age). To highlight the broad age distribution in the current sample, Figure 1 shows a histogram of the participant’s ages in years. Average years of education was 14.2 ± 2.0 years. The mean estimated premorbid IQ of the sample, measured with the Wechsler test of adult reading (WTAR; Wechsler, 2001), was 104.4 ± 10.5 . Thus, participants showed on average normal intellectual functioning. Compared to the known demographics of the original sample in Esterman et al. (2013), there are some differences which would be expected from a Veteran population compared to a predominantly college-based sample. The current sample was significantly more biased toward male participants, 37.5% vs. 93.6% male: $\chi^2 = 43.63$, $p < 0.0001$, and a Mann-Whitney U test showed that on average the Veteran sample was older, 24 vs. 32 years of age: $U=426.5$, $p < 0.001$. As years of education were not recorded in the original sample, no comparison could be calculated though it is likely that a sample recruited from college campuses may have a higher number of years of education on average than the 14 years average in the current sample. We additionally note that our sample of Veterans, while limited in terms of gender, includes a range of individuals in terms of background demographics and clinical issues. While beyond the scope of the current paper, this sample includes Veterans with a history of mild TBI (e.g. concussions), and a range of potential

clinical symptoms including anxiety, PTSD, and depression. We include additional clinical information about our sample in Supplementary Figure 1 for the interested reader.

Finally, we compared the motion parameters across the two studies using Mann-Whitney U tests looking at the maximum absolute displacement across the entire run. These results show no overall difference across the six motion parameters in our current sample to those measured in Esterman et al. (2013). A trend towards smaller yaw rotations was observed in the present sample (Current Sample = 0.5479° ; Esterman 2013 sample = 0.6162° ; $U = 795$, $p = 0.058$) while no other motion directions showed any difference ($p > 0.24$ for all). The means for the six motion parameters across the two groups is shown in Supplementary Figure 2.

2.2 Behavioral Paradigm and Stimuli

For this study, a single 8-minute run of the gradCPT was completed, following the methodology of Esterman et al. (2013) with one exception. As in the original study, the gradCPT stimuli consisted of 20 round, grayscale photographs of mountain and city scenes, with 10 from each category. On each trial, a random scene was chosen for presentation with 90% probability that the chosen scene would be a city scene and 10% probability that the scene would be a mountain scene, with the constraint that identical images could not be chosen on consecutive trials. Using linear interpolation, scene images gradually faded from one to the next over the course of 800ms, for a total of 600 trial images over the course of the 8-minute run. For the task, participants were required to press a button when a city scene was shown and withhold responses when mountain scenes were shown (i.e., go/no-go task). The task instructions emphasized response accuracy without reference to speed. However, as a new image replaced the previous image every 800ms, there was an implicit response deadline in the task. Participants were given a practice session prior to scanning where they were familiarized with each of the 20 scene images and given 1–2 minutes of practice completing the task. In contrast to the original study (Esterman et al., 2013), which used a goggle system for stimulus presentation, images in the current study were viewed from a mirror in the scanner projected via a back-projection screen.

2.3 MRI Acquisition and Processing

Scanning was completed at the Neuroimaging Research for Veterans (NeRVe) center at the VA Boston Healthcare hospital on a 3T Siemens MAGNETOM Trio system. During the first part of the scanner session two anatomical magnetization prepared rapid gradient-echo (MP-RAGE) structural scans were obtained with a 12-channel head coil. These MP-RAGE T-1 scans were acquired with the following parameters: repetition time (TR) = 2530ms, echo time (TE) = 3.32ms, flip angle = 7° , acquisition matrix = $256 \times 256 \times 176$, voxel size = 1mm^3 . Anatomical scans were inspected at acquisition for motion artifacts and repeated if necessary. Following acquisition these scans were averaged to increase signal-to-noise ratio. All structural images were then processed using standard FreeSurfer and Analysis of Functional Imaging (AFNI) pipelines (Cox, 1996; Fischl, Sereno, & Dale, 1999; Fischl et al., 2004).

The functional run was collected using a 32-channel head coil and one whole-brain echo-planar T2*-weighted sequence. The scanning parameters for the functional scan were as follows: TR = 2000ms, TE = 30ms, flip angle = 90°, 248 volumes, acquisition matrix = 64 × 64, in-plane resolution = 3.0 × 3.0 mm², slice thickness = 3.75mm. Following acquisition, the functional scan was processed using AFNI and custom written routines in Matlab (Mathworks Inc., Natick, MA). Preprocessing steps included slice-time correction, motion correction using a 6-parameter, rigid body, least-squares alignment procedure, spatial smoothing with a 6-mm FWHM Gaussian kernel, automated co-registration and normalization of anatomical and functional volumes to Talairach space, and scaling of functional dataset values to percent signal change using the equation $x' = 100 * (x - x_0) / x_0$, where x_0 was the mean value of the run. During preprocessing, automated segmentation algorithms generated three masks from the Talairached anatomical volume. These included masks covering grey matter, white matter, and cerebral spinal fluid (CSF). Average time series from the functional scan were extracted from eroded white matter and cerebral spinal fluid masks to use as nuisance regressors.

2.4 Behavioral Analyses

2.4.1 Reaction time and accuracy—Analysis of behavioral performance on the gradCPT has been described in detail elsewhere (Esterman et al., 2013; Fortenbaugh et al., 2015). Briefly, response times to each trial were determined using an iterative algorithm that assigned button presses to individual trials. Reaction times were calculated relative to the beginning of each image onset. Thus, a reaction time of 800ms would indicate that the current image was 100% coherent while shorter reaction times indicate that the current image was still in the process of transitioning in from the previous image. After the response time algorithm was run, reaction time and performance variables were calculated. These include: mean reaction time, reaction time variability (defined using the coefficient of variation (CV), or the standard deviation of the reaction time divided by the mean reaction time for that participant), commission error (CE) rate (the number of target mountain scenes a participant pressed to), and omission error (OE) rate (the number of city scenes a participant failed to press to). Using standard signal detection analysis, the commission and omission error rates were then used to calculate d' and criterion scores, where d' reflects the ability of participants to discriminate between city and mountain scenes and criterion reflects the strategy used by participants, or the willingness to press the response button in the case of uncertainty. Standard procedures were used to correct for cases where hit rates were 100% or false alarm rates were 0%, with one-half error deducted or added, based on the total number of target or non-targets presented in the run, respectively. A confirmatory factor analysis was performed, following up on the exploratory factor analyses that were completed on the four primary behavioral variables examined in web-based versions of the gradCPT (Fortenbaugh et al., 2015)—namely, the two reaction time measures, d' , and criterion. The confirmatory factor analysis was run in R using a maximum likelihood estimator (Lavaan R package; Rosseel, 2012), specifying a 2-factor model of ability (CV and d') and strategy (RT and criterion).

2.4.2 Vigilance—To investigate potential vigilance decrements over the course of the 8-minute run, linear changes in performance were assessed for each of the six behavioral

measures: mean reaction time, reaction time variability, commission error rate, omission error rate, d' , and criterion. For this analysis, the data was divided into four 2-minute quartiles. Mean performance was then assessed for each quartile and a linear regression was calculated to determine the slope parameter for each participant separately. One-sample t -tests were conducted to determine if the slopes differed significantly from zero.

2.4.3 In the zone vs. out of the zone—Following the analysis first outlined in Esterman et al. (2013), this analysis inferred instantaneous attentional state by using trial-by-trial variations in reaction time to calculate the variance time course (VTC). VTCs were computed for each participant using the >500 correct responses to the non-target city scenes. First, reaction times were z -transformed to normalize values within participants and the absolute value of the z -scores was calculated so that higher values indicated greater deviations from the mean, including both very slow and very fast reaction times, while lower values indicated reaction times closer to the mean of the run. Values for trials without responses (omission errors and correct omissions to target mountain scenes) were linearly interpolated from the reaction times of the two surrounding trials. A smoothed VTC was then computed using a Gaussian kernel of 9 trials full-width at half-maximum (FWHM), integrating information from the surrounding 20 trials with a weighted average. While the VTC is a continuous, within-subject measure of variability that is agnostic of when participants are in any particular attentional state, splitting the time series up into high and low variability trials allows for accuracy comparisons across a participant's relatively more or less stable periods of performance. From the smoothed VTC, a median split was used to divide performance on each trial into low and high-variability bins. Based on previous work, these 4-minute periods are referred to as being “in the zone” and “out of the zone”, respectively. While there are multiple ways that can be used to split performance (e.g., relative to the group-level mean variability, or splitting two or more contiguous intervals of trials into groups), the trial-by-trial assignment into low/high variability bins was used in order to follow previous methodology and provide a within-subject measurement of each participant's state relative to their overall performance (Esterman et al., 2013).

2.5 Neuroimaging Analyses

2.5.1 Event Related Activity—For each participant, functional data was submitted to a hierarchical general linear model (GLM) analysis that included two levels. The first level analysis was used to regress motion and nuisance parameters from each voxel's time series. The first-level GLM included regressors associated with the six motion parameters and two additional nuisance regressors: cerebral spinal fluid (CSF) and white matter time series. Additionally, linear, quadratic, and cubic trends were modeled. The GLM analysis also included censoring of time-points around abrupt movements. The time-points where motion exceeded 0.5mm as well as the TR immediately following the movement were censored (ignored) in the GLM analysis (see Supplementary Figure 2 for individual participant censoring proportions). This affected the estimation of regressor beta-values but did not delete the time points from the residual time series. The residual time series from the first-level GLM were then submitted to a second-level GLM that modeled the stimulus events, again using time-point censoring. Stimulus events, including correct omissions, commission errors, and omission errors were modeled as impulse events in the same GLM. Correct

commissions (accurately responding to non-target city scenes) were not explicitly modeled in this GLM due to their high frequency. Regression coefficients for each of the three event types were compiled across participants and tested via voxel-wise one-sample t-tests. The correct omission versus commission error contrast was tested with a paired-sample t-test. Whole-brain statistical maps were corrected for multiple comparisons using the new and more conservative ex-Gaussian voxel-cluster Monte-Carlo-type α simulation rather than the previous standard Gaussian model in AFNI (Cox, Chen, Glen, Reynolds, & Taylor, 2017). First, the AFNI 3dFWHMx function was run with the spatial autocorrelation function (ACF) option in order to estimate the spatial smoothness of the data with a mixed Gaussian plus mono-exponential model to generate random noise fields. The estimated parameters for this model were then used with the 3dClustSim command and ACF option to estimate the minimum cluster sizes needed to reach statistical significance. Across all three event types, the correction omission/commission error contrast, and the two whole-brain VTC analyses below, we choose the most conservative cluster-size threshold corrected $p < 0.05$ for all analyses, at a nominal $p=0.01$. Significant clusters were 81 voxels.

2.5.2 Lapse Precursors—To model pre-trial activity in response to target mountain scenes when participants either correctly withheld responses (correct omissions) or incorrectly pressed (commission errors), two different analytic approaches were utilized. For both analysis, the residual time series from the first-level GLM which did not model events were used. As AfNI zeros out but does not delete censored time points in the residuals, after extraction of the time series the time points censored in the GLM analyses were converted to NaN in Matlab for all following analyses and were thus again excluded from these analyses. First, we replicated the approach utilized in Esterman et al. (2013). Using the same, independently-defined, regions of interest (ROIs) from Esterman et al. (2013), three ROIs were assessed: the parahippocampal place area (PPA), dorsal attention network (DAN), and the default mode network (DMN). In the second approach, we looked at pretrial activity using a whole-brain voxel-level analysis. Average residual time series were extracted for each of the three ROIs in the first approach, while individual residual time series were extracted for each voxel in the second approach. Using an iterative algorithm, linear time interpolation was conducted to estimate the BOLD response at each image transition (rate = 0.8 sec), using only the nearest TRs for estimation. For the ROI analysis, the pretrial period was defined as the average activation of the two trials immediately preceding a target mountain onset (−1.6 sec to −0.8 sec). For the second approach, in order to accommodate the increased noise moving from an ROI to voxel-level time series analysis, we increased the temporal averaging window to seven trials (~3TRs) that occurred in the window −4.8sec to 0.0sec before target onset. To assess statistical significance in this second analysis, a permutation-based Monte Carlo approach was used. The correct omission and commission error labels were randomly shuffled for each participant. A new subtraction map was calculated and the maximum cluster size observed for positive and negative differences in this shuffled map was determined. This shuffling process was repeated 5,000 times and the distribution of maximum positive and negative cluster sizes for the shuffled subtraction maps was calculated. From these distributions, the 250th largest cluster size (top 5th percentile) was recorded separately for positive and negative cluster distributions. Taking the larger of these two cluster sizes resulted in a 353-voxel cluster threshold. Note we did not perform the

lapse precursor analysis split by in/out of the zone given the small number of commission error trials when participants were in the zone within a single run (i.e. 19 participants did not have a single commission error when they were in the zone and the average number of in the zone commission errors was ~4 trials).

In addition to calculating the mean pre-trial BOLD signal change, we extended analyses from previous studies using the gradCPT by examining the temporal dynamics across our three a priori ROI regions (PPA, DAN, and DMN) to determine the extent to which patterns of pretrial activity across these three regions is predictive of upcoming lapses or successes in sustained attention. For this analyses, we included only those participants who had 3 or more commission error trials (N=135) in order to ensure enough time points were available for the slope analysis. Of the 5 participants excluded from this analysis, one participant had no commission errors and thus no data available in this category, and two participants each had only one or two commission errors across the entire run. Using a design similar to that applied by Thompson et al. (2013), we assessed temporal dynamics in a 12.8 second window (16 trials) centered 6.4 seconds prior to the onset of a mountain stimulus (see Figure 6A). First, we extracted the time points in the 12.8 sec window for each trial and modeled temporal changes in BOLD activity using linear regression to obtain a slope value. For each participant, separate trial-based slope values were calculated for correct omission and commission error trials in each of the three ROIs. The mean slope was then calculated across all the trials in each category. We then calculated the mean slope values prior to commission errors or correct omissions for each of the three ROIs individually and compared differences in slope values using a 3 (ROI) \times 2 (Event Type) repeated-measures ANOVA and paired t-tests.

2.5.3 Variance Time Course—To determine which voxels had time series that co-varied with fluctuations in attention state, a slightly different approach from the original study by Esterman et al. (2013) was used. We note, however, that this time-delayed approach is reported in the Supplementary Materials of that paper and compared with the main analytic approach. As in Esterman et al. (2013), the residuals from the second-level event-related GLM analysis outlined above were used. Given the delayed hemodynamic response, the smoothed VTC time series for each participant was then shifted by 6 seconds and whole brain correlations were calculated using 3dRegAna in AFNI. Spearman rho calculations were used rather than Pearson's r as the question of interest was which brain regions show monotonic trends in fluctuations in BOLD activity with changes in reaction time variability, and there was no a priori reason to assume these fluctuations would be linearly related to reaction time. Additionally, inspection of individual participant's VTC time series showed that data points are generally not normally distributed given the lower limit on reaction times variability (i.e., absolute z-scores) leading to positive skewed distributions. Thus, non-parametric correlations are a more appropriate statistic in this type of analysis. The resulting correlation coefficients were converted to Fisher z scores and a one-sample t-test was calculated on normalized correlation coefficients. As with the event-related analyses above, results of the analysis were thresholded using the cluster-corrected thresholds for $p < 0.05$ (nominal $p = 0.01$; cluster size = 81 voxels).

2.5.4. Variance Time Course and Individual Differences—While the degree to which multiple brain regions track the variance time course has been assessed in previous studies, to date no work has examined the extent to which the degree of coupling is related to or predictive of overall performance. To determine whether the degree of coupling between BOLD signal and reaction time variability is meaningful for an individual's overall performance, two additional analyses were completed that examined inter-individual differences in overall performance on the gradCPT. First, we tested whether the z-scores from the above whole-brain VTC analysis were correlated with participant's overall accuracy on the task using their d' score. For this analysis, at every voxel the VTC regression z-scores were correlated with d' scores across participants. As with the other whole-brain analyses above, results of the analysis were thresholded using the cluster-corrected thresholds for $p < 0.05$ (nominal $p = 0.01$; cluster size = 81 voxels).

In our second analysis, using our PPA, DMN, and DAN ROIs, we tested whether consideration of the degree to which each region tracks a participant's VTC and the degree to which regions are coupled with each other over the course of the run can predict individual differences in overall performance, again using d' as our dependent variable. For this, the 10 nodes of the ROIs were used (PPA = 2 nodes, left and right; DAN = 4 nodes, left and right intraparietal sulcus (IPS) and frontal eye fields (FEF); DMN = 4 nodes, bilateral anterior medial prefrontal cortex (AMPFC), bilateral posterior cingulate cortex (PCC), and left and right lateral parietal cortex; see Supplementary Figure 4). We first calculated the correlation (Pearson's r) between the VTC and the average time series from each of the 10 nodes as well as the correlation between each of the node pairs for each participant. This gave a total of 55 connections/VTC-brain correlations which were used in a multiple linear regression model using a leave-one-subject-out (LOSO) cross-validation procedure to predict d' (Esterman, Tamber-Rosenau, Chiu, & Yantis, 2010; Fortenbaugh, Corbo, et al., 2017). For the LOSO procedure, every participant was left out of the training dataset once. In each of these iterations, a linear regression model was built from N features across the 139 participants in the training dataset, where N was varied from 1 to 55 (all possible) features. Feature selection was done by finding the N features with the greatest correlation between coupling and d' in the training set. Thus, for a given feature set size, the features included in the model could vary across participants. Once the features were selected the beta weights from the regression model were used to predict the d' score of the participant left out. After all 140 iterations for a given feature set size, the predicted and measured d' scores across all 140 participants were correlated with each other to assess the quality of the model.

3. Results

3.1. Behavioral Results

3.1.1 Overall performance—While the primary focus of the current paper was to investigate whether the neuroimaging findings of Esterman et al. (2013) replicated in the current sample, it was first important to consider whether our novel sample of Veterans performed behaviorally similar to the previous sample of 16 young, healthy non-Veterans as any performance differences would be expected to impact the related BOLD signal. Table 1

summarizes the behavioral performance of the 140 Veterans in the current sample. For comparison, Table 1 also shows the mean performance from the 1st run of the experiment in Esterman et al. (2013) as participants in this study completed multiple runs of the task. Given the significantly larger sample size in the present study, data from Esterman et al. (2013) were compared to the present study using Mann-Whitney U tests. As seen in Table 1, for all performance measures including both reaction time and accuracy, at the group level, performance did not differ significantly in the present sample to that observed over the 1st run of the task in Esterman et al. (2013).

3.1.2 Time-on-task effects—The original study by Esterman et al. (2013) also considered changes in performance across each run, to investigate whether vigilance decrements could be observed in how participants completed this challenging sustained attention task. As seen in Table 1, a similar pattern of change in performance across the 8 minutes of task was observed. For all four of the primary performance measures considered in Esterman et al. (2013), namely, mean reaction time, reaction time variability (CV), commission errors, and omission errors, significant changes in performance were observed as indicated by the one-sample t-tests comparing the group slope values to a hypothetical mean of zero. The present study also included two additional performance measures, d' and criterion, which represent discrimination ability and strategy on task, respectively. Interestingly, while discrimination ability was seen to significantly decrease over the course of the run, consistent with a vigilance decrement, we found no change in the strategy used by participants. Comparing the slopes to those observed in the original sample (Esterman et al., 2013), the only behavioral difference was in the slope of the mean reaction times, with participants in the present sample speeding up their reaction times over the course of the run while participants in the original sample tended to slow down. Importantly, however, for all performance measures related to accuracy, strategy, or fluctuations in reaction time, changes in performance across time were equivalent across the two samples.

3.1.3 Relationship between performance variables—In the original study by Esterman et al. (2013), a strong correlation was observed across participants between reaction time variability (CV) and lapse rate, defined by the number of commission errors. In Fortenbaugh et al. (2015) this relationship was further investigated using factor analyses on the behavioral performance of over 10,000 participants across the world who completed an online version of the gradCPT task. Results of these analyses indicated two latent factors in performance on the gradCPT task. The first factor, which we have labeled the ability factor, was driven by reaction time variability and discrimination ability, measured using d' to account for both commission and omission error rates. The second factor, the strategy factor, was driven by mean reaction time and criterion values, and reflected the approach used by participants to complete the task. As seen in Figure 2 and Table 2, the results of the present study replicated the findings found in Fortenbaugh et al. (2015), with strong correlations observed across the ability variables, d' and reaction time variability and the strategy variables, criterion and mean reaction time. While all correlations show a significant relationship, further analyses were completed to compare the strength of the correlations taking into account the covariance matrix given that the correlations were from the same participants and the measures dependent (Steiger, 1980). As seen in Table 2, results showed

that the relationship between d' to reaction variability is significantly stronger than the relationship between d' and mean reaction time. Similarly, the relationship between criterion and mean reaction time is significantly stronger than the relationship between criterion and reaction time variability (see Table 2). To assess if the same latent variables were observed in the present sample as previous studies (Fortenbaugh et al., 2015), a confirmatory factor analysis was run using the proposed two factor model (Table 3). We found that this model fit the data sufficiently well with, according to current standards (Hu & Bentler, 1999; Schreiber, Nora, Stage, Barlow, & King, 2006), four out of five criteria indicating a good fit (see Table 3).

3.1.4 In the zone vs. out of the zone—Using the VTC time series to calculate stable reaction time periods (“in-the-zone”) and unstable periods (“out-of-the-zone”), we examined whether behavioral performance differs across these periods of time. As seen in Figure 3, consistent with the study by Esterman et al. (2013), paired-sample t-tests show that participants made more errors when they were out of the zone than in the zone (commission error rate: $t(139) = 12.01, p < 0.0001$; omission error rate: $t(139) = 9.14, p < 0.0001$), while no difference was seen in the mean reaction time across the two epochs, $t(139) = 1.10, p = 0.27$. We further compared our signal detection measures across in-the-zone and out-of-the-zone epochs. Results showed a decrease in discrimination ability (d') when participants were out of the zone, $t(139) = 17.47, p < 0.0001$, while no change in response strategy were observed in the criterion measure, $t(139) = 1.295, p = 0.197$.

3.2. fMRI Results

3.2.1 Event-Related Activation Analyses—Analysis of event-related activity focused on three types of events: 1) correct omissions, where participants correctly withheld responses to the rare mountain scenes, 2) commission errors, where participants failed to withhold a response to mountain scenes, and 3) omission errors, or rare trials where participants failed to press to a city scene. Figure 4 shows maps of areas with significant event-related BOLD activity to these three types of events and the contrast between correct omissions and commission errors (see Supplementary Tables 1–3 for detailed cluster information). As seen in Figure 4a, the most frequent type of event, correct omissions, was associated with widespread increases in task positive networks associated with attention, including the dorsal attention, fronto-parietal, and salience networks (Yeo et al., 2011). This widespread positive activity also covers regions associated with the vigilance attention network (Langner & Eickoff, 2013), specifically the pre-supplementary motor area (pre-SMA), inferior frontal gyrus, insula, lateral prefrontal cortex, temporal-parietal junction, intraparietal sulcus, middle occipital gyrus, temporal occipital junction, and thalamus. In contrast, decreases in activity are seen in default network regions, including the anterior medial prefrontal cortex (amPFC), posterior cingulate cortex (PCC), left lateral parietal cortex. Within visual cortex, a division was seen with medial visual cortex showing a decrease in activity while lateral and ventral visual cortex showed an increase in activity. These activations were likely the results of a combination of target detection, cognitive control and response inhibition necessary to withhold a response to target mountain stimuli.

Commission errors, or failures to effectively detect and/or withhold responses to rare mountain trials (Figure 4b), were associated with increased activity in the ventral attention and salience networks, also referred to as the cingulo-opercular network, thought to be engaged in error monitoring, attentional reconfiguration, and maintaining/refreshing tonic alertness, task control and goals (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Sadaghiani & D'Esposito, 2015), including the bilateral insula and fronto-operculum, thalamus, dorsal anterior cingulate cortex including pre-SMA, and anterior prefrontal cortex. Figure 4d shows the correct omission vs. commission error map, with positive values showing areas with greater activity following correct omissions while negative values show relatively higher activity following a commission error. Lateral visual and parietal regions, including the IPS which is known to contain topographic maps of spatial attention (M. A. Silver & Kastner, 2009; M. A. Silver, Ress, & Heeger, 2005), showed greater responses on correct omission trials, potentially reflecting greater visual attention to stimuli during these trials. On the other hand, greater activation in dorsal ACC and insular cortex for commission errors may have reflected error-related activation or processing.

The second type of error that participants made in this type of not-X CPT task was omission errors, where participants fail to respond to frequent city images. Task-evoked activity to omission errors (Figure 4c) was almost entirely overlapping with the evoked responses to commission errors and correct omissions, with increased activity seen in the pre-SMA region, thalamus, inferior parietal lobules, middle frontal gyrus, and insula, and decreased activity in the posterior cingulate. This may have reflected some aspect of response inhibition, independent of task-relevance (CO and OE), as well as error-related activation, independent of stimulus type (CE and OE).

3.2.2 Precursors of Attention Lapses—While the previous analysis examined the evoked responses to correct omissions, commission errors, and omission errors using a standard GLM approach assuming a canonical hemodynamic response, in our next analysis we examined BOLD activity related to these events in the time leading up to the target mountain trials (correct omission and commission errors). Following the approach outlined in Esterman et al. (2013), we first examined pre-trial activity averaged across the two trials immediately preceding the beginning of each mountain trial (−1.6 to −0.8 secs from trial beginning). Note that one participant made no commission errors across the run and was therefore excluded from the following analyses. Figure 5a shows the mean deviation in BOLD signal for the PPA, DAN, and DMN ROIs across this window. Paired-sample t-tests showed only a partial replication of the results from Esterman et al. (2013). BOLD activity was significantly lower in the PPA and higher in the DMN prior to lapse trials when participants made commission errors compared to trials where participants correctly withheld responses (PPA: $t(138) = -2.25$, $p = 0.026$; DMN: $t(138) = 3.05$, $p = 0.003$). In contrast to Esterman et al. (2013), while numerically the DAN showed lower BOLD activity in lapse trials compared to correct response trials, no significant difference was observed across commission error and correct omission trials, $t(138) = -0.88$, $p = 0.382$. To further explore the consistency across studies, we plot in Table 4 the mean BOLD activity contrast (correct omissions - commission errors) along with the 95% confidence intervals from the original Esterman et al. (2013) dataset, along with the measured contrast from the current

data. As can be seen in Table 4, not only was no significant difference observed between commission errors and correct omission trials for the DAN ROI in the present dataset, but the difference across these two types of trials fell outside the 95% confidence interval of the significant contrast measured in Esterman et al. (2013). Thus, while the magnitude of the difference was small within the DAN ROI in the original study, the observed difference fell outside the expected range, indicating that the current results do not replicate those of the original study.

Extending this analysis beyond Esterman et al. (2013), we examined differences between precursors of correct omissions vs. commission errors at the whole brain voxel-wise level (Figure 5b; see Supplementary Table 4 for detailed cluster information). We found significant clusters in ventromedial PFC regions consistent with the ROI-level analysis showing higher activity in the DMN before a commission error. Interestingly, we also found a cluster in somatomotor cortex in the right hemisphere extending to left somatomotor cortex, suggesting greater activity in the task-relevant motor system or fine motor preparation signals could lead to errors of commission (right-handed response). Conversely, left insula, right ventral visual cortex, and bilateral thalamus exhibited greater activity before correctly withholding, regions consistent with response control and visual processing respectively. Overall these results were consistent with the ROI analyses and prior work, but do not provide support for DAN as a precursor to response accuracy, as was found in the original report.

3.2.3 Temporal Dynamics of Pretrial Activity—While the previous analysis examined mean activity levels in the time just prior to a target stimulus onset, examining mean activity within a small window of time provides just a snapshot of overall activity and may not fully capture the temporal dynamics that lead to subsequent lapses in performance. While there are many potential ways to model temporal dynamics, in the following analysis we examined temporal changes in pretrial BOLD activity using a simple measurement of change over time: linear slope analysis (i.e., moving from a point estimate to a line). Figure 6a shows a schematic of the procedure used, which involved estimating slope parameters across the 16 trials (12.8sec window) preceding a target onset and examining differences in slope patterns across event type (correct omissions and omission errors) for the PPA, DAN, and DMN ROIs. Comparing the mean slope parameter estimates across participants with a 2×3 repeated-measures ANOVA (see Figure 6b), we found no main effect of event type, $F(1,134) = 2.45, p = 0.120$ but a main effect of ROI, $F(2,268) = 6.34, p = 0.002$. Importantly, however, an Event Type × ROI interaction was observed, $F(2,268) = 3.50, p = 0.032$, with lower slopes prior to commission errors than correct omissions in the task-positive PPA and DAN ROIs while higher slopes prior to commission errors were seen in the DMN. Follow-up paired sample t-tests showed that slopes significantly differed across correction omission and commission error trials in the task-positive ROIs (PPA: $t(134) = -2.16, p = 0.032$; DAN: $t(134) = -2.14, p = 0.034$), while no significant difference was found in the DMN ROI, $t(134) = 0.87, p = 0.387$.

3.2.4 Reaction Time Stability—To examine the relationship between trial-by-trial fluctuations in reaction time stability and BOLD activity, whole-brain correlations were

calculated using each participant's VTC time series. For this correlation, the VTC time series were down-sampled into TR-space and shifted by 3 TRs (6 sec) to account for the hemodynamic delay. Correlations were calculated using Spearman Rho and the resulting correlation coefficients were Fisher transformed. Figure 7 shows the resulting VTC thresholded map (see Supplementary Table 5 for detailed cluster information). Consistent with previous studies (Esterman et al., 2013; Esterman, Rosenberg, et al., 2014; Kucyi et al., 2017), several networks showed significant correlations with reaction time variability. Within the DMN, bilateral posterior cingulate and ventromedial prefrontal cortex, and left lateral parietal cortex all showed negative correlations with the VTC time series, indicating higher levels of activity during moments of relative stability (i.e. lower reaction time variability). Negative correlations were also seen extending from the posterior cingulate cortex inferiorly into the medial occipital cortex. In contrast, positive correlations were observed in several task-positive regions including supplementary motor association area, lateral and ventral occipital cortex, bilateral inferior frontal gyrus and insula, frontal eye fields, and temporo-parietal junction. These areas corresponded to areas typically engaged during sustained attention tasks, including the dorsal and ventral attention networks and the salience network.

3.2.5 Reaction Time Stability and Overall Performance—While multiple task-positive and task-negative regions were found to track fluctuations in reaction time variability, and this replicates several previous studies (Esterman et al., 2013; Esterman et al., 2017; Esterman, Rosenberg, et al., 2014; Kucyi et al., 2016), these relationships are counter to other characterizations of these networks (for review see: Fortenbaugh, DeGutis, et al., 2017). For example, DMN was more active “in the zone”, while task positive networks were more active “out of the zone” (see Discussion). Given this, our next analysis went a step further in asking whether this coupling of BOLD activity and response variability is predictive of overall behavioral performance at the individual level. For this, the Fisher z-transformed correlation coefficients from each participant's VTC analysis were correlated on a voxel-by-voxel basis with overall discrimination ability (d') across participants. Results of this analysis (Figure 8; see Supplementary Table 6 for detailed cluster information) showed that the degree to which the posterior cingulate cortex (overlapping with DMN) tracked reaction time variability on a moment-to-moment basis predicted overall task performance (d') across participants. Greater negative coupling (lower variability and higher PCC activation) in an individual was associated with more accurate performance. A similar relationship was found for a region in the left superior frontal gyrus which also overlaps with the DMN (Yeo et al., 2011). In contrast, regions in right inferior parietal/temporal-parietal junction (overlapping with VAN) and ventral visual cortex showed the opposite pattern—greater positive couple (higher variability and higher temporal-parietal junction activation) was associated with more accurate performance. Overall, these results suggest that if anything, these relationships between ongoing variability and brain activity (Figure 7) were generally reflective of better performance at the individual level.

3.2.6 Predicting Individual Differences in Performance Using VTC and Network-Level—In our final analysis we utilized the three networks/ROI groupings from the Esterman et al. (2013) study to determine whether the degree of coupling between PPA,

DAN, and DMN regions with reaction time stability and overall connectivity across network regions can be used to predict individual differences in overall performance. Across the VTC and 10 nodes comprising the PPA, DAN, and DMN ROIs (see Methods Section 2.5.4. *Variance Time Course and Individual Differences* and Supplementary Figure 4 for details), we tested the ability of 1–55 features to predict d' using a leave-one-subject-out linear regression approach. Figure 9a shows the correlation coefficient between the measured and predicted d' values across all 140 participants as a function of the number of features used in the linear regression models. As seen in Figure 9a, the best prediction ability was seen with the 18-Feature model, which showed a significant positive correlation between the measured and predicted d' values, $r = 0.402$, $p < 0.0001$ (see Figure 9b). For the 18-Feature model, a total of 25 unique connections were selected across all 140 participant folds and 11/25 connections were selected on 140/140 participant folds. Of the 11 connections selected in every fold, the predominant groupings included VTC-DMN connections and DMN-DAN connections (see Figure 9c). The pattern held when all connections selected were grouped according to ROIs (see Table 5). Across the 25 connections, 10 included DAN-DMN connections while 9 included VTC-ROI connections (4 VTC-DMN, 4 VTC-DAN, 1 VTC-PPA). Collectively, these results suggest that important information regarding attentional stability and overall performance is carried in both the degree to which the DAN and DMN track reaction time stability, or the VTC, and that additional unique information is carried in the degree to which DAN and DMN activity is correlated with each other.

4. Discussion

The results of the present study provide both an important robust replication of findings from the original gradCPT study (Esterman et al., 2013), as well as an extension by relating the strength of VTC-brain coupling to individual differences in performance and pretrial temporal dynamics to lapse likelihood. In the original gradCPT study (Esterman et al., 2013), 16 young, healthy volunteers completed multiple runs of the task while in the scanner. In contrast, the present study utilized 140 Veteran participants who each completed one run of the task. While Veterans can be considered a special population, with higher prevalence rates of both neurological and psychiatric illnesses than the general population (e.g., mild traumatic brain injury and posttraumatic stress disorder; see below), it is important to note that at the group level, the performance of the Veteran participants in the present study was statistically identical to that of the participants in the original study for all behavioral measures with the exception of one: the change in mean reaction time over the course of the run. Additional behavioral analyses comparing the relationship across four of the primary behavioral measures using a confirmatory factor analysis (Table 3) provided support for the same latent factors related to ability and strategy on the task as found in a large web-based sample of over 10,000 participants between the ages of 10–70 years old who completed a shorter version of the gradCPT task (Fortenbaugh et al., 2015). Collectively then, the behavioral results of the present sample suggest that performance of this unique sample are both valid and representative of what one would expect at the group-level from the general population.

More importantly, perhaps, given recent concerns that the field of neuroimaging faces a replication crisis (Poldrack et al., 2017), the present neuroimaging results provide a robust

replication of the relationship between fluctuations in response variability and ongoing brain activity. Specifically, the whole-brain VTC correlation (Figure 7) replicated and extended the results of Esterman et al. (2013), and corroborated similar analyses in subsequent papers (Esterman et al., 2017; Esterman, Rosenberg, et al., 2014; Kucyi et al., 2016). Namely, we found that regions in the default mode network exhibited greater activity during low variability in-the-zone periods, as does the putamen. On the other hand, task-positive regions in the dorsal and ventral attention network exhibited greater activity during highly variable, out-of-the-zone periods. Our results, along with these previous studies, suggest that in-the-zone performance is accomplished with either less attentional resources (i.e., more effortlessly), or alternatively, that attentional resources are engaged with greater efficiency and precision. Studies have supported the latter; task-irrelevant stimuli are processed with greater depth during these in-the-zone periods, akin to lower perceptual load, or more efficient task-related processing (Esterman, Rosenberg, et al., 2014). Similarly, we have found that inhibitory TMS (1 Hz) to the right frontal eye field impaired more optimal periods of performance (in the zone), but not less optimal (out of the zone) periods of performance (Esterman et al., 2015). This suggests that despite lower overall activity levels in dorsal and ventral attention regions while in the zone, these task-positive regions may be more critical during these periods of consistent and accurate performance. In the domain of attentional control over the motor system, Kucyi et al. (2017) found that stable and accurate periods of rhythmic finger tapping was also associated with less activation in task-positive regions and more activation in DMN regions, again consistent with neural efficiency and suggestive that the relationship between attentional stability and neural coupling is not restricted to visual attention. On the other hand, recent evidence suggests that DMN activity may be greater during periods of high predictability, suggesting that in the zone periods could be akin to a more “autopilot” mode (Vatansever, Menon, & Stamatakis, 2017).

While the variability-brain correlation replicated prior group-level results, we also took advantage of the large sample size to ascertain whether this coupling between ongoing brain activity and behavioral variability was associated with individual differences in performance. While the VTC has been shown behaviorally to track relative performance within individuals (i.e., better accuracy in versus out of the zone), the extent to which individual differences in VTC coupling may relate to overall performance has not been examined before. If, on the one hand, the VTC correlations in the brain were stronger in those with worse performance, it would suggest that this coupling is maladaptive. If, however, this brain-behavior coupling is stronger in better performers, it suggests that this correlation is adaptive. Our results from the whole-brain VTC coupling correlation analyses (Figure 8) suggest the latter, such that subcomponents of the DMN show stronger negative coupling (more activity in the zone) in more accurate performers. Similarly, the right temporal-parietal junction of the ventral attention network exhibited stronger positive coupling with variability (more activity out of the zone) in more accurate performers. These individual-differences results, while not as extensive as the group VTC result, suggest that the coupling observed between fluctuations of attention and ongoing activity in large-scale brain networks is adaptive. Additionally, focusing just on the PPA, DAN, and DMN networks, our cross-validation linear regression analysis demonstrated that the degree of VTC coupling within the nodes of the DAN and DMN in particular, in addition to coupling across the DAN

and DMN nodes, contains discriminative information that can be used to predict individual differences in overall performance. These results provide further support that the extent to which regions couple with fluctuations in reaction time variability, while intrinsically a within-subject measure of attentional state, also contains information regarding inter-individual differences in sustained attention performance. In the original study, it was posited that the relative balance of DMN and DAN activity may play an important role in determining attentional state at any given moment, though no interactive effects were tested at that time (Esterman et al., 2013). The theory that the interplay between task-positive and task-negative regions, and in particular the interplay between the DAN and DMN, rather than the activity in one region alone, may have an important role in cognitive functioning is receiving increasing interest (Avelar-Periera, Backman, Wahlin, Nyberg, & Salami, 2017; de Pasquale, Corbetta, Betti, & Penna, in press; Gao & Lin, 2012; Kucyi et al., 2017; Thompson et al., 2013; Vatansever et al., 2017). The fact that the degree of DMN and DAN connectivity was able to provide additional information beyond just the VTC-DMN and VTC-DAN features in the regression models supports this assertion. Given the limited number of a priori regions used in the present analysis, however, future work is needed to determine how widespread or localized this connectivity-derived information is across the brain.

The current study also sought to replicate and extend previous analyses on the neural precursors of attentional lapses, defined as errors of commission. Broadly, the results across the two pretrial analyses show that changes in BOLD activity leading up to a target event can provide information regarding future performance, or lapse likelihood. Consistent with the original report and others (Esterman et al., 2013; Esterman, Rosenberg, et al., 2014; Thompson et al., 2013), mean activity in stimulus specific PPA was higher just prior to correct vs. lapses trials, while DMN showed the opposite pattern, greater mean activity preceding lapses vs. correct trials. Extending this analysis, examining the temporal dynamics across correct omission and commission error trials showed different patterns of activity across the three ROIs in the 12.8 seconds prior to a target onset. While PPA and DAN showed negative slopes prior to a lapse trial, with BOLD activity decreasing, results showed if anything an increase in DMN BOLD activity, though the contrast across correct omission and commission error slopes was not significant for DMN. As seen in Figure 6, the differences in slopes was primarily seen for lapse trials across the three ROIs with little to no difference in activity patterns across the PPA, DMN, and DAN ROIs (slopes ≈ 0) prior to correct omission trials. These results suggest that PPA and DAN activity represents task-related stimulus processing that enables successful performance, and that when this activity wanes over many seconds, performance can suffer. On the other hand, it suggests that DMN activity may reflect task-unrelated activity, such as mind wandering (Christoff et al., 2016; Garrison et al., 2013; Konishi, McLaren, Engen, & Smallwood, 2015; Kucyi et al., 2016; Raichle et al., 2001). This is consistent with several studies, including Kucyi et al. (2016), which used thought probes to examine DMN activity in relation to both mind wandering and variability. Interestingly, they also found that while DMN activity was greater during mind wandering, it was also greater during periods of low variability (i.e., in the zone, see previous paragraph). This was the case despite overall greater variability during mind wandering. Additionally, there was no anatomical segregation between DMN voxels that

tracked behavioral variability versus those that tracked mind wandering. Rather, independent and additive influences of attentional state (in/out of the zone) and mind-wandering were assessed using linear mixed-effects models on the average DMN time series, and were again found when tested on DMN sub-regions parcellating the DMN into 7 or 54 sub-regions. These results are also consistent with findings from a recent study assessing cognitive flexibility (Vatansever et al., 2017) which suggests a role for the DMN in automated information processing related to the performance of external task goals (i.e., not just mind wandering). Together, the current results, along with Kucyi et al. (2017) and Vatansever et al. (2017), argue against a unitary model of DMN and suggest that multiple, independent contributions of the DMN to cognition exist (e.g., in relation to behavioral variability, accuracy, and mind wandering).

In contrast to the original report, we did not replicate the significant precursor effects in the DAN (see Figure 5 and Table 4), though the slope analysis showed a significant difference in the direction of BOLD activity slopes prior to attentional lapses versus correct omissions. In the original study (Esterman et al., 2013), higher mean levels of activity in the DAN was associated with subsequent correct vs. lapses trials. In the current data, this effect did not reach significance and the observed difference fell outside of the 95% confidence interval from Esterman et al. (2013). There are several possible reasons why this component of the study failed to replicate. One possible explanation regards the role of motivation. We previously found that task positive networks, including DAN, exhibited greater activity before correct trials only when participants were motivated (with monetary rewards) relative to an unrewarded condition (Esterman et al., 2017). This led to the conclusion that motivation is associated with a more proactive strategy (Botvinick, Braver, Barch, Carter, & Cohen, 2001), such that participants actively maintain activity in task positive regions in order to maximize success. Thus, such motivational differences could explain the weaker nature of the current results, which may have varied more in this sample because participants completed the gradCPT task at the end of a long day of assessments. Conversely, it may be that prior to target onsets pretrial effects are weaker in DAN than PPA, with differences evolving and becoming more prominent after target onsets. The fact that significant slope differences were observed in the pretrial window could be seen as supporting this argument, with the dynamic changes in the DAN perhaps more important than the final value just prior to target onset. Along these lines, it may be that given the small sample size in the original study, the mean pretrial differences observed just before target onset in the DAN are in fact unreliable. Given recent work suggesting that changes in short-window temporal dynamics may be related to cognitive functioning (Kucyi et al., 2017; Thompson et al., 2013), and our potentially conflicting finding regarding changes in DAN activity leading to correct omissions and commission errors, additional work will be needed to decide between these competing explanations.

Finally, using a whole brain approach, we found that activity in the motor cortex was also associated with lapses of attention. We speculate that greater activity in the contralateral motor system could indicate a more pre-potent response set, increasing the likelihood of a failure to inhibit a response in the upcoming target/mountain trial. Peak activation in this cluster was observed in the right (ipsilateral) motor cortex extending across into left motor cortex (see Figure 5 and Supplementary Table 4). While ipsilateral, and specifically the

right, motor cortex has been associated with fine finger movements (Chen, Gerloff, Hallett, & Cohen, 1997), it is unclear from the present data the extent to which such ipsilateral signals may reflect general motor planning signals. Together, these results suggest that there are heterogeneous neural, and perhaps psychological causes of attentional lapses (e.g., motor, mind wandering, effort). Future work should further examine whether individual and clinical differences in types of lapses exist and whether they can be independently predicted and even modulated. For example, TMS directed toward DMN regions could be used to reduce lapses due to mind wandering, while stimulation of visual cortices could attenuate lapses due to visual distraction. Further, trial-to-trial prediction, or even real-time fMRI could take advantage of a heterogeneous model of attentional lapses to enhance accuracy (deBettencourt, Cohen, Lee, Norman, & Turk-Browne, 2015).

In addition to the analyses of the original report, we also examined and contrasted the evoked responses to the different stimulus event types. Target/mountain events evoked broad activation in task-positive and lateral visual regions, as well as a deactivation in DMN and medial visual regions, regardless of whether participants correctly withheld their response (Figure 4a and 4b). Interestingly, correct omissions evoked relatively greater responses in lateral visual and dorsal attention, implicating enhanced visual processing and attentional control in the execution of response inhibition (Figure 3d). In contrast, errors of commission evoked relatively greater responses in salience network regions including dorsal anterior cingulate (dACC) and insular cortex, regions implicated in error processing and task reconfiguration. Finally, rare errors of omission evoked responses in regions that were associated with commission errors (insular and dACC) as well as others associated with correct omissions (dorsal parietal/prefrontal). In general, all task events evoked highly overlapping patterns of activation that call into question pure regional specificity of functions such as inhibition or error monitoring (Aron, Robbins, & Poldrack, 2004; Botvinick, Cohen, & Carter, 2004; Kerns et al., 2004; Swick & Chatham, 2014).

There are several limitations in the current study. One important limitation is the low number of female participants, with 94% of our sample being male. While this is representative of the U.S. Veteran population, important gender differences have previously been reported in the sustained attention literature (Blatter et al., 2006; Riley et al., 2016). For healthy samples, this concern is somewhat reduced by the fact that our results largely replicate those of Esterman et al. (2013), in which the sample was 62.5% women. However, from a clinical perspective, understanding how sustained attention varies as a function of gender is an important issue that requires further study to better understand individual differences and provide a framework for understanding changes in sustained attention in clinical populations that may include more women than men. Additionally, the current sample included all Veterans enrolled into the TRACTS cohort who were able to complete an MRI scanning session. As no additional screening or inclusion/exclusion criteria were applied, within the sample as a whole, there are multiple sub-groups of participants (e.g., Veterans with mTBI, clinical symptoms of anxiety, depression, or PTSD). The goal of the present study was to take a group-level approach to replication and the fact that clinical sub-samples were present in the current group of participants represents more of a limitation for the individual differences analysis than the group-level evoked events, VTC, and pretrial analyses. However, as with any selective sample, the present data must therefore be approached with

some caution in terms of generalizing the results to the general population or other potential clinical samples. Finally, while the gradCPT paradigm is similar to many other continuous performance tasks that utilize a go/no-go response paradigm (Conners, 1994; Robertson et al., 1997), the ability to completely segregate event-related effects attributable to attentional lapses vs. response preparation/inhibition is reduced. While there are separate issues that may be faced using alternative response paradigms (e.g., 2AFC), future research examining differences across paradigms with differing response patterns may help to further dissociate the underlying cognitive processes that contribute to the evoked response patterns observed.

In spite of these limitations, this robust replication has many implications for future research focused on clinical populations with attentional dysfunction. Importantly, this study indicates that previous findings using this neurocognitive measure of sustained attention will likely generalize to other samples with greater diversity or unique characteristics in terms of demographics or clinical pathology. Although beyond the scope of this study, we are hopeful this task will help elucidate mechanisms of attentional impairments across a wide range of clinical pathologies that are common in this population and others, including PTSD, mTBI, and substance abuse (DeGutis et al., 2015; Swick, Honzel, Larsen, & Ashley, 2013; Swick, Honzel, Larsen, Ashley, & Justus, 2012). Additionally, while the neuroimaging results of the current study focused on group-level results, future studies focused on individual differences within the general population may provide an opportunity to assess the neural basis of normal variations in sustained attention ability. Functional connectivity during the task may also add additional mechanistic understanding of how sustained attention can reveal clinically relevant brain dysfunction. Research has already demonstrated that connectivity during the gradCPT is associated with early life trauma and ADHD symptoms (Fortenbaugh, Corbo, et al., 2017; Rosenberg et al., 2016). Given that deficits in sustained attention are pervasive across a wide range of clinical populations (for review, see: Fortenbaugh, DeGutis, et al., 2017), the present results may provide a basis for exploring cognitive and neural dysfunction across a wide range of patient groups.

In conclusion, the current study robustly replicates the majority of behavioral and neuroimaging markers of sustained attention and fluctuations in accuracy and variability over time. As the current sample represents a more demographically and clinically diverse group than most basic fMRI research, the present results further underscore the validity of the gradCPT and supports its generalizability to other populations. As this task has already been used by multiple research groups and cited hundreds of times, the current study provides valuable assurance to future investigators who contemplate this methodology to study sustained attention. More broadly, we hope this work serves to highlight the importance and utility of replication in the fMRI community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was support by the US Department of Veterans Affairs through the Translational Research Center for TBI and Stress Disorders (TRACTS), a VA Rehabilitation Research & Development Traumatic Brain Injury

National Research Center B9254-C, a Career Development award from the Department of Veterans Affairs Clinical Sciences Research and Development Service (IK2CX000706-01A2) to M.S.E., a National Institute of Health NCCIH grant (R21 AT009430-01) to R.M., and an Advanced Geriatric Fellowship from the Department of Veterans Affairs to F.C.F. The contents of this article do not represent the views of the Department of Veterans Affairs or the U.S. government.

References

- Altpeter E, Mackeben M, Trauzettel-Klosinski S. The importance of sustained attention for patients with maculopathies. *Vision Research*. 2000; 40(10–12):1539–1547. DOI: 10.1016/S0042-6989(00)00059-6 [PubMed: 10788657]
- Andrews - Hanna JR, Smallwood J, Spreng RN. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*. 2014; 1316(1):29–52. [PubMed: 24502540]
- Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*. 2004; 8(4):170–177. [PubMed: 15050513]
- Auerbach RP, Kim JC, Chango JM, Spiro WJ, Cha C, Gold J, ... Nock MK. Adolescent nonsuicidal self-injury: Examining the role of child abuse, comorbidity, and disinhibition. *Psychiatry Research*. 2014; 220(1–2):579–584. [PubMed: 25095754]
- Avelar-Periera B, Backman L, Wahlin A, Nyberg L, Salami A. Age-related differences in dynamic interactions among default mode, frontoparietal control, and dorsal attention networks during resting-state and inference resolution. *Frontiers in Aging Neuroscience*. 2017; 9(152):1–15. DOI: 10.3389/fnagi.2017.00152 [PubMed: 28174533]
- Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*. 1997; 121(1):65–94. [PubMed: 9000892]
- Berardi A, Parasuraman R, Haxby J. Overall vigilance and sustained attention decrements in healthy aging. *Experimental Aging Research*. 2001; 27:19–39. [PubMed: 11205528]
- Blatter K, Graw P, Münch M, Knoblauch V, Wirz-Justice A, Cajochen C. Gender and age differences in psychomotor vigilance performance under differential sleep pressure conditions. *Behavioural Brain Research*. 2006; 168(2):312–317. [PubMed: 16386807]
- Boekel W, Wagenmakers EJ, Belay L, Verhagen J, Brown S, Forstmann BU. A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*. 2015; 66:115–133. [PubMed: 25684445]
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. *Psychological Review*. 2001; 108(3):624–652. [PubMed: 11488380]
- Botvinick MM, Cohen JD, Carter CS. Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*. 2004; 8(12):539–546. [PubMed: 15556023]
- Broyd SJ, Demanuele C, Debener S, Helps SK, James CJ, Sonuga-Barke EJ. Default-mode brain dysfunction in mental disorders: a systematic review. *Neuroscience & Biobehavioral Reviews*. 2009; 33(3):279–296. [PubMed: 18824195]
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews: Neuroscience*. 2013:14.
- Carriere JSA, Cheyne JA, Solman G, JF, Smilek D. Age trends for failures of sustained attention. *Psychology and Aging*. 2010; 25(3):569–574. [PubMed: 20677878]
- Chen R, Gerloff C, Hallett M, Cohen LG. Involvement of the ipsilateral motor cortex in finger movements of different complexities. *Annals of Neurology*. 1997; 41(2):247–254. [PubMed: 9029074]
- Christoff K, Irving ZC, Fox KC, Spreng RN, Andrews-Hanna JR. Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience*. 2016; (17):718–731. [PubMed: 27654862]
- Clark L, Iversen SD, Goodwin GM. Sustained attention deficit in bipolar disorder. *The British Journal of Psychiatry*. 2002; 180(4):313–319. DOI: 10.1192/bjp.180.4.313 [PubMed: 11925353]

- Clayton MS, Yeung N, Cohen Kadosh R. The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*. 2015; 19(4):188–195. DOI: 10.1016/j.tics.2015.02.004 [PubMed: 25765608]
- Conners, C. *The Conners Continuous Performance Test*. Toronto, Canada: Multi-Health Systems; 1994.
- Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*. 2002; 3(3):201–215. DOI: 10.1038/nrn755 [PubMed: 11994752]
- Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*. 1996; 29(3):162–173. DOI: 10.1006/cbmr.1996.0014 [PubMed: 8812068]
- Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. FMRI Clustering in AFNI: False-Positive Rates Redux. *Brain Connectivity*. 2017; 7(3):152–171. [PubMed: 28398812]
- de Pasquale F, Corbetta M, Betti V, Penna SD. Cortical cores in network dynamics. *NeuroImage*. :1–13. in press. DOI: 10.116/j.neuroimage.2017.09.063
- deBettencourt M, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*. 2015; 18(3):470–475. [PubMed: 25664913]
- DeGutis J, Esterman M, McCulloch B, Rosenblatt A, Milberg W, McGlinchey R. Posttraumatic Psychological Symptoms are Associated with Reduced Inhibitory Control, not General Executive Dysfunction. *Journal of the International Neuropsychological Society*. 2015; 21(05):342–352. [PubMed: 26126233]
- Dosenbach NU, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*. 2008; 12(3):99–105. [PubMed: 18262825]
- Esterman M, Grosso M, Liu G, Mitko A, Morris R, DeGutis J. Anticipation of monetary reward can attenuate the vigilance decrement. *PLoS One*. 2016; 11(7):e0159741. 0159741-0159719. [PubMed: 27472785]
- Esterman M, Liu G, Okabe H, Reagan A, Thai M, DeGutis J. Frontal eye field involvement in sustaining visual attention: Evidence from transcranial magnetic stimulation. *NeuroImage*. 2015; 111:542–548. [PubMed: 25655445]
- Esterman M, Noonan SK, Rosenberg MD, DeGutis J. In the Zone or Zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*. 2013; 23(11):2712–2723. DOI: 10.1093/cercor/bhs261 [PubMed: 22941724]
- Esterman M, Poole V, Liu G, DeGutis J. Modulating Reward Induces Differential Neurocognitive Approaches to Sustained Attention. *Cerebral Cortex*. 2017; 27(8):4022–4032. DOI: 10.1093/cercor/bhw214 [PubMed: 27473320]
- Esterman M, Reagan A, Liu G, Turner C, DeGutis J. Reward reveals dissociable aspects of sustained attention. *Journal of Experimental Psychology: General*. 2014; 143(6):2287–2295. DOI: 10.1037/xge0000019 [PubMed: 25313950]
- Esterman M, Rosenberg MD, Noonan SK. Intrinsic fluctuations in sustained attention and distractor processing. *The Journal of Neuroscience*. 2014; 34(5):1724–1730. [PubMed: 24478354]
- Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S. Avoiding non-independence in fMRI data analysis: Leave one subject out. *NeuroImage*. 2010; 50(2):572–576. DOI: 10.1016/j.neuroimage.2009.10.092 [PubMed: 20006712]
- Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 1999; 9(2):195–207. DOI: 10.1006/nimg.1998.0396 [PubMed: 9931269]
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, ... Dale AM. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*. 2004; 14(1):11–22. [PubMed: 14654453]
- Forster S, Nunez Elizalde AO, Castle E, Bishop SJ. Unraveling the Anxious Mind: Anxiety, Worry, and Frontal Engagement in Sustained Attention Versus Off-Task Processing. *Cerebral Cortex*. 2015; 25(3):609–618. DOI: 10.1093/cercor/bht248 [PubMed: 24062316]
- Fortenbaugh FC, Corbo V, Poole V, McGlinchey R, Milberg W, Salat D, ... Esterman M. Interpersonal early-life trauma alters amygdala connectivity and sustained attention performance. *Brain and Behavior*. 2017; 7(5):e00684. 00681-00616. doi: 10.1002/brb3.684 [PubMed: 28523226]

- Fortenbaugh FC, DeGutis J, Esterman M. Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*. 2017; 1396:70–91. [PubMed: 28260249]
- Fortenbaugh FC, DeGutis J, Germine L, Wilmer JB, Grosso M, Russo K, Esterman M. Sustained Attention Across the Life Span in a Sample of 10,000: Dissociating Ability and Strategy. *Psychological Science*. 2015; 26(9):1497–1510. DOI: 10.1177/0956797615594896 [PubMed: 26253551]
- Fortenbaugh FC, Robertson LC, Esterman M. Changes in the distribution of sustained attention alter the perceived structure of visual space. *Vision Research*. 2017; 131:26–36. DOI: 10.1016/j.visres.2016.12.002 [PubMed: 28025055]
- Gao W, Lin W. Frontal parietal control network regulates the anti-correlated default and dorsal attention networks. *Human Brain Mapping*. 2012; 33(1):192–202. [PubMed: 21391263]
- Garrison KA, Scheinost D, Worhunsky PD, Elwafi HM, Thornhill TA, Thompson E, ... Hampson M. Real-time fMRI links subjective experience with brain activity during focused attention. *NeuroImage*. 2013; 81:110–118. [PubMed: 23684866]
- Greicius MD, Krasnow B, Reiss AL, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*. 2003; 100(1):253–258.
- Helton WS, Russell PN. Working memory load and the vigilance decrement. *Experimental Brain Research*. 2011; 212(3):429–437. [PubMed: 21643711]
- Hu, Lt, Bentler, PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*. 1999; 6(1):1–55.
- Johnson BP, Pinar A, Fornito A, Nandam LS, Hester R, Bellgrove MA. Left anterior cingulate activity predicts intra-individual reaction time variability in healthy adults. *Neuropsychologia*. 2015; 72:22–26. [PubMed: 25791710]
- Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. *Science*. 2004; 303(5660):1023–1026. [PubMed: 14963333]
- Konishi M, McLaren DG, Engen H, Smallwood J. Shaped by the past: the default mode network supports cognition that is independent of immediate perceptual input. *PLoS One*. 2015; 10(6):e0132209. [PubMed: 26125559]
- Kucyi A, Esterman M, Riley CS, Valera EM. Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences*. 2016; 113(48):13899–13904.
- Kucyi A, Hove MJ, Esterman M, Hutchison RM, Valera EM. Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral Cortex*. 2017; 27(3):1831–1840. DOI: 10.1093/cercor/bhw029 [PubMed: 26874182]
- Langner R, Eickoff SB. Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*. 2013; 139(4):870–900. [PubMed: 23163491]
- Lawrence NS, Ross TJ, Hoffmann R, Garavan H, Stein EA. Multiple Neuronal Networks Mediate Sustained Attention. *Journal of Cognitive Neuroscience*. 2003; 15(7):1028–1038. DOI: 10.1162/089892903770007416 [PubMed: 14614813]
- Levy F. The development of sustained attention (vigilance) and inhibition in children: Some normative data. *Journal of Child Psychology and Psychiatry*. 1980; 21(1):77–84. DOI: 10.1111/j.1469-7610.1980.tb00018.x [PubMed: 7358807]
- Lippa SM, Fonda JR, Fortier CB, MAA, AK, Milberg WP, McGlinchey R. Deployment-Related Psychiatric and Behavioral Conditions and Their Association with Functional Disability in OEF/OIF/OND Veterans. *Journal of Traumatic Stress*. 2015; 28:155–184.
- Mackworth NH. The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*. 1948; 1(1):6–21. DOI: 10.1080/17470214808416738

- Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN. Wandering minds: the default network and stimulus-independent thought. *Science*. 2007; 315(5810):393–395. [PubMed: 17234951]
- McGlinchey RE, Milberg WP, Fonda JR, Fortier CB. A methodology for assessing deployment trauma and its consequences in OEF/OIF/OND veterans: The TRACTS longitudinal prospective cohort study. *International journal of methods in psychiatric research*, e1556. 2017; :1–15. DOI: 10.1002/mpr.1556
- Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015; 349(6251):aac4716–aac4716. DOI: 10.1126/science.aac4716 [PubMed: 26315443]
- Park M, Hood MM, Shah RC, Fogg LF, Wyatt JK. Sleepiness, parkinsonian features and sustained attention in mild Alzheimer’s disease. *Age and Ageing*. 2012; 41(6):765–770. DOI: 10.1093/ageing/afs084 [PubMed: 22743151]
- Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, ... Yarkoni T. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*. 2017; 18:115–126. DOI: 10.1038/nrn.2016.167 [PubMed: 28053326]
- Posner MI, Peterson SE. The attention system of the human brain. *Annual Review of Neuroscience*. 1990; 13:25–42.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proceedings of the National Academy of Sciences*. 2001; 98(2):676–682.
- Riley E, Esterman M, Fortenbaugh FC, DeGutis J. Time-of-day variation in sustained attentional control. *Chronobiology International*. :1–9. in press.
- Riley E, Okabe H, Germine L, Wilmer J, Esterman M, DeGutis J. Gender Differences in Sustained Attentional Control Relate to Gender Inequality across Countries. *PLoS One*. 2016; 11(11):e0165100. [PubMed: 27802294]
- Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. ‘Oops!’: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*. 1997; 35(6):747–758. DOI: 10.1016/S0028-3932(97)00015-8 [PubMed: 9204482]
- Rosenberg MD, Finn ES, Constable RT, Chun MM. Predicting moment-to-moment attentional state. *NeuroImage*. 2015; 114:249–256. [PubMed: 25800207]
- Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, Chun MM. A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*. 2016; 19(1):165–171. DOI: 10.1038/nn.4179 [PubMed: 26595653]
- Rosenberg MD, Noonan SK, DeGutis J, Esterman M. Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics*. 2013; 75(3):426–439.
- Rosseel, Y. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). Ghent, Belgium: Ghent University; 2012.
- Sadaghiani S, D’Esposito M. Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*. 2015; 25(9):2763–2773. [PubMed: 24770711]
- Sarter M, Givens B, Bruno JP. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews*. 2001; 35:146–160. [PubMed: 11336780]
- Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*. 2006; 99(6): 323–338.
- Schretlen DJ, Shapiro AM. A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry*. 2003; 15(4):341–349. [PubMed: 15276955]
- Shalev L, Ben-Simon A, Mevorach C, Cohen Y, Tsai Y. Conjunctive Continuous Performance Task (CCPT)—A pure measure of sustained attention. *Neuropsychologia*. 2011; 49(9):2584–2591. [PubMed: 21609725]
- Silver H, Feldman P. Evidence for sustained attention and working memory in schizophrenia sharing a common mechanism. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 2005; 17(3): 391–398. DOI: 10.1176/jnp.17.3.391 [PubMed: 16179663]
- Silver MA, Kastner S. Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*. 2009; 13(11):488–495. DOI: 10.1016/j.tics.2009.08.005 [PubMed: 19758835]

- Silver MA, Ress D, Heeger DJ. Topographic Maps of Visual Spatial Attention in Human Parietal Cortex. *J Neurophysiol.* 2005; 94(2):1358–1371. DOI: 10.1152/jn.01316.2004 [PubMed: 15817643]
- Slovaryp L, Azuma T, LaPointe L. The effect of traumatic brain injury on sustained attention and working memory. *Brain injury.* 2012; 26(1):48–57. [PubMed: 22149444]
- Staub B, Doignon-Camus N, Bacon E, Bonnefond A. Investigating sustained attention ability in the elderly by using two different approaches: Inhibiting ongoing behavior versus responding on rare occasions. *Acta Psychologica.* 2014; 146(1):51–57. [PubMed: 24378237]
- Staub B, Doignon-Camus N, Després O, Bonnefond A. Sustained attention in the elderly: What do we know and what does it tell us about cognitive aging? *Ageing Research Reviews.* 2013; 12:459–468. DOI: 10.1016/j.arr.2012.12.001 [PubMed: 23261761]
- Steiger JH. Tests for comparing elements of a correlation matrix. *Psychological Bulletin.* 1980; 87:245–251.
- Swick D, Chatham CH. Ten years of inhibition revisited. *Frontiers in Human Neuroscience.* 2014; 8:329. [PubMed: 24904369]
- Swick D, Honzel N, Larsen J, Ashley V. Increased response variability as a marker of executive dysfunction in veterans with post-traumatic stress disorder. *Neuropsychologia.* 2013; 51(14):3033–3040. [PubMed: 24157540]
- Swick D, Honzel N, Larsen J, Ashley V, Justus T. Impaired response inhibition in veterans with post-traumatic stress disorder and mild traumatic brain injury. *Journal of the International Neuropsychological Society.* 2012; 18(05):917–926. [PubMed: 22595028]
- Temple JG, Warm JS, Dember WN, Jones KS, LaGrange CM, Matthews G. The Effects of Signal Salience and Caffeine on Performance, Workload, and Stress in an Abbreviated Vigilance Task. *Human Factors.* 2000; 42(2):183–194. [PubMed: 11022879]
- Thompson GJ, Magnuson ME, Merritt MD, Schwarb H, Pan WJ, McKinley A, ... Keilholz SD. Short - time windows of correlation between large - scale functional brain networks predict vigilance intraindividually and interindividually. *Human Brain Mapping.* 2013; 34(12):3280–3298. [PubMed: 22736565]
- Van Vleet TM, DeGutis JM. Cross-training in hemispatial neglect: Auditory sustained attention training ameliorates visual attention deficits. *Cortex.* 2013; 49(3):679–690. [PubMed: 22578712]
- Vatansever D, Menon DK, Stamatakis EA. Default mode contributions to automated information processing. *Proceedings of the National Academy of Sciences.* 2017; 114(48):12821–12826. DOI: 10.1073/pnas.1210521114
- Wechsler, D. Wechsler Test of Adult Reading: WTAR. Psychological Corporation; 2001.
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, ... Buckner RL. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology.* 2011; 106(3):1125–1165. DOI: 10.1152/jn.00338.2011 [PubMed: 21653723]

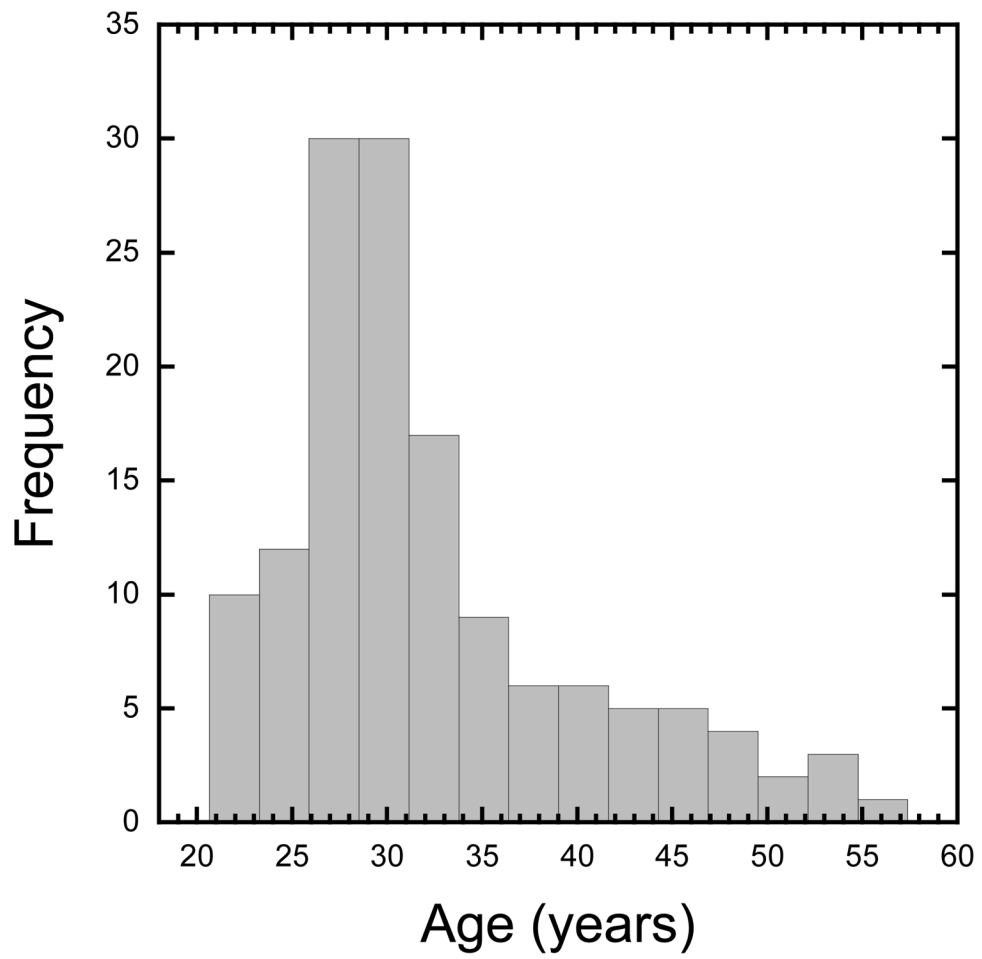


Figure 1.
Histogram showing the ages of participants in years.

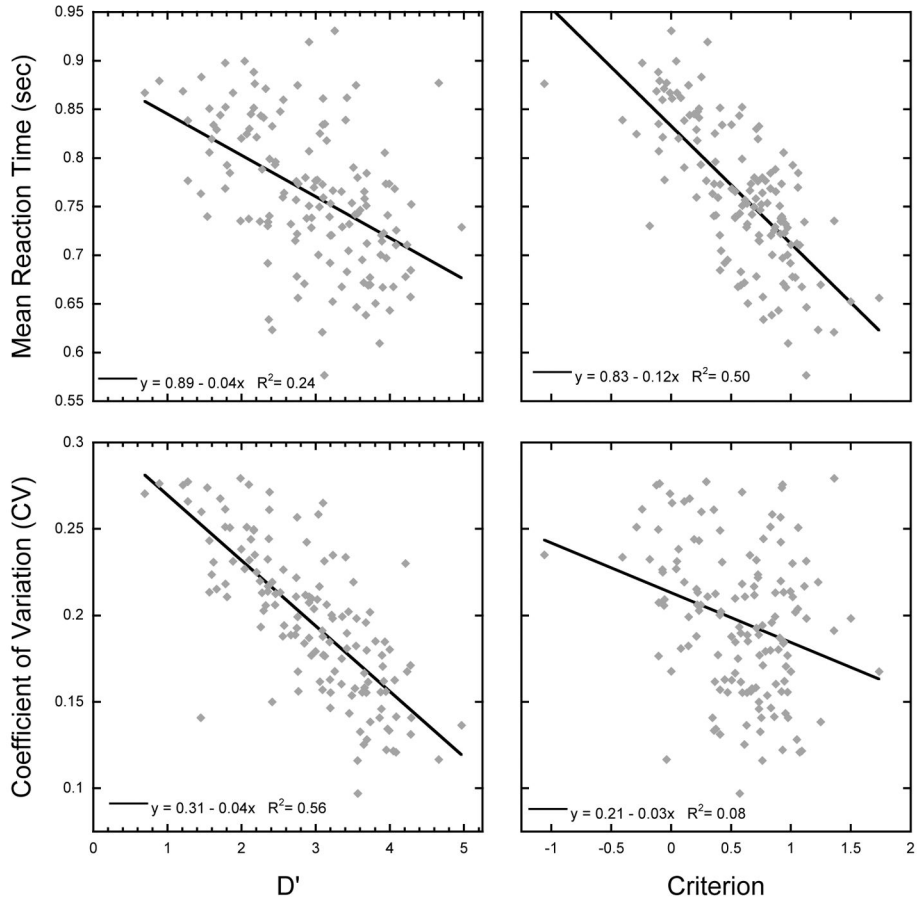


Figure 2. Behavioral Results

Scatterplot showing the relationship across the four primary behavioral measures on the gradCPT task identified in Fortenbaugh et al. (2015): mean reaction time, reaction time variability (CV), d' , and criterion. The graph shows strong correlations within the behavioral measures related to performance (d' and CV) and the measures related to strategy (criterion and RT). Each diamond represents a single participant and the black line shows the regression line.

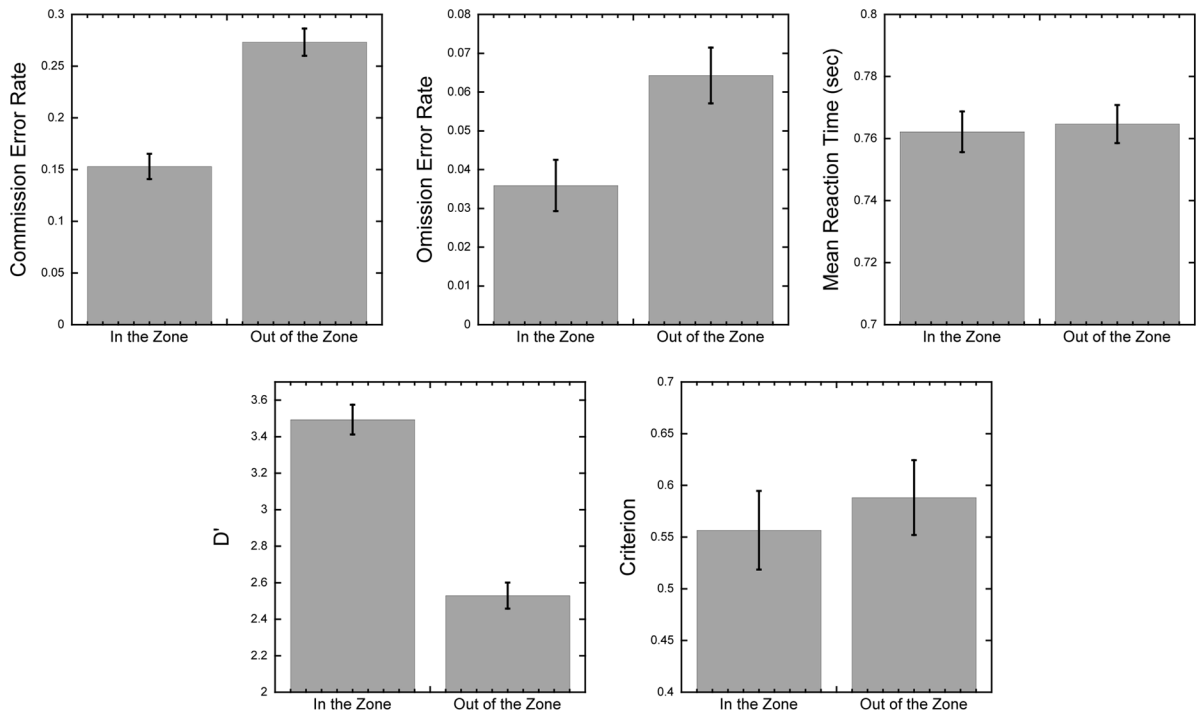
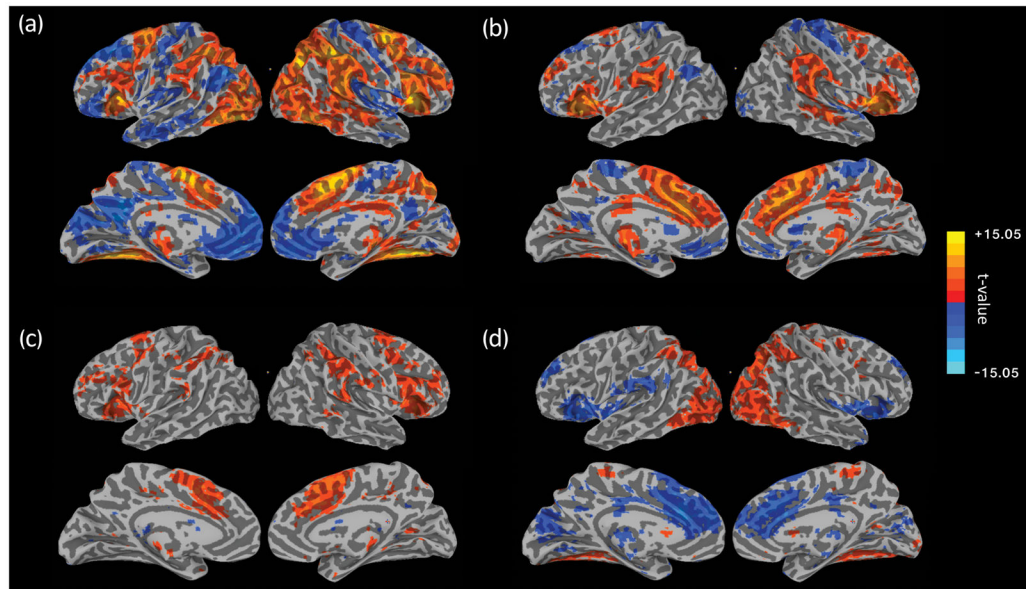


Figure 3. Behavioral Results

Bar graphs showing mean behavioral performance differences when participants are in-the-zone versus out-of-the-zone for the behavioral performance measures. Error bars represent ± 1 S.E.M.

**Figure 4. Evoked Activity Results**

Event evoked activity results from GLM analysis for (a) Correct Omissions, (b) Commission Errors, (c) Omission Errors, and (d) Correct Omissions – Commission Error contrast. All maps show T-statistics and are displayed after correction for multiple comparisons (corrected $p < 0.05$; nominal $p < 0.01$, cluster size > 81 voxels).

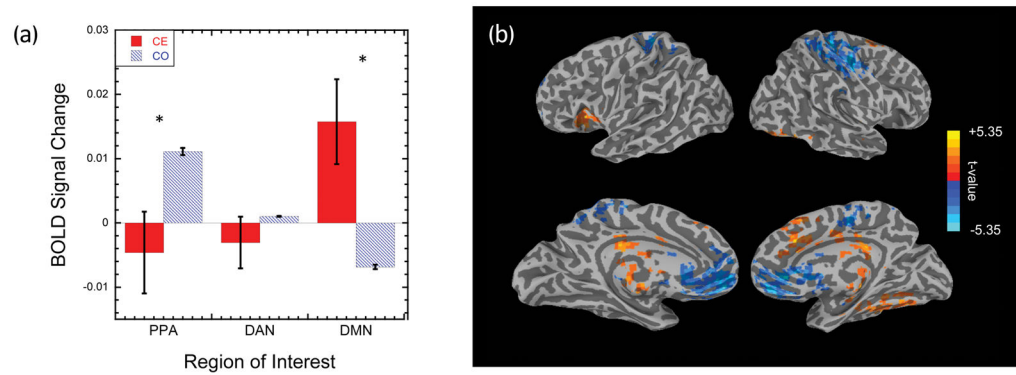


Figure 5. Pre-trial Activity Results

(a) The left panel shows the results from ROI-level analysis. The average activation level for Commission Errors (CE) and Correct Omissions (CO) across the -1.6sec to -0.8sec window prior to target onset are shown for the PPA, DAN, and DMN ROIs. Error bars show ± 1 S.E.M. Significant differences across CO and CE pretrial activity are shown with * for $p < 0.05$. (b) The right panel shows the results from the whole-brain voxel-level analysis on the CO-CE contrast activity averaged across the -4.8 to 0.0sec window prior to target onset. This map shows the T-statistic thresholded after correction for multiple comparisons (Monte-Carlo $p < 0.05$, cluster size > 353 voxels).

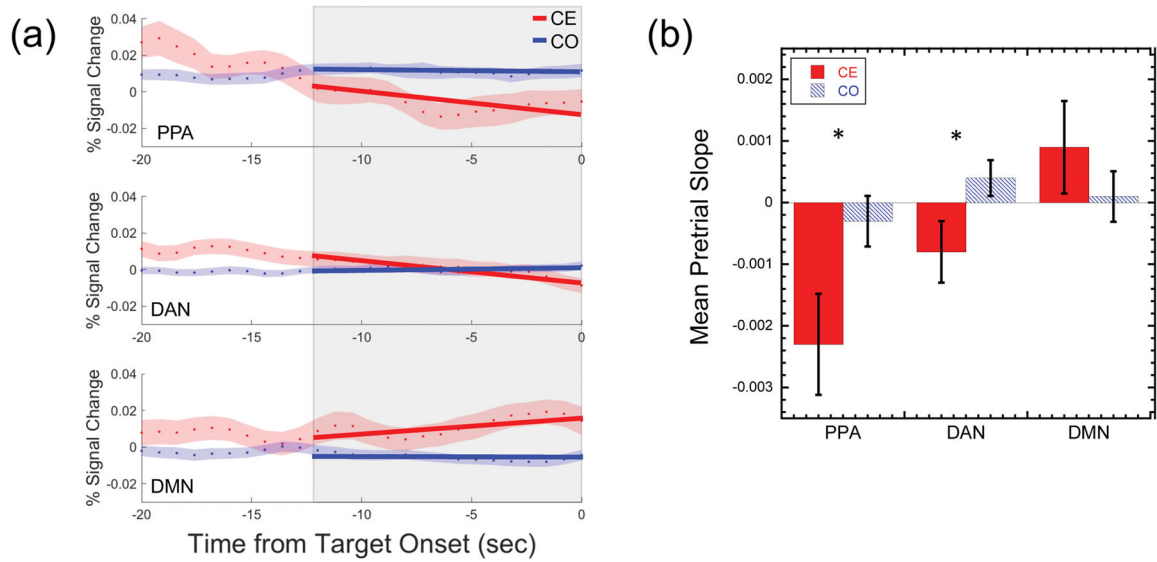


Figure 6. Network-Level Trends Prior to Attentional Lapses

(a) The panels on the left side of the figure illustrate the method used to assess dynamic pretrial activity in individual participants. The mean time series in the window -20sec to 0sec prior to target scene onsets is shown for Correct Omission (CO) and Commission Error (CE) trials for the PPA, DAN, and DMN ROIs separately. The blue and red dots show the mean across trials with the red and blue shaded regions shows ± 1 S.E.M. The shaded gray box shows the 12.8sec (16 trials) interval of interest just prior to target onset used in the current analysis. Linear regression was applied to the 16 time points in each trial to extract a slope parameter (examples shown by red and blue lines). (b) Group level bar graph showing the mean slope across participants for CO and CE trials as a function of ROI (PPA, DAN, and DMN). Error bars show ± 1 S.E.M. Significant differences in slopes across CO and CE trials are shown with * for $p < 0.05$.

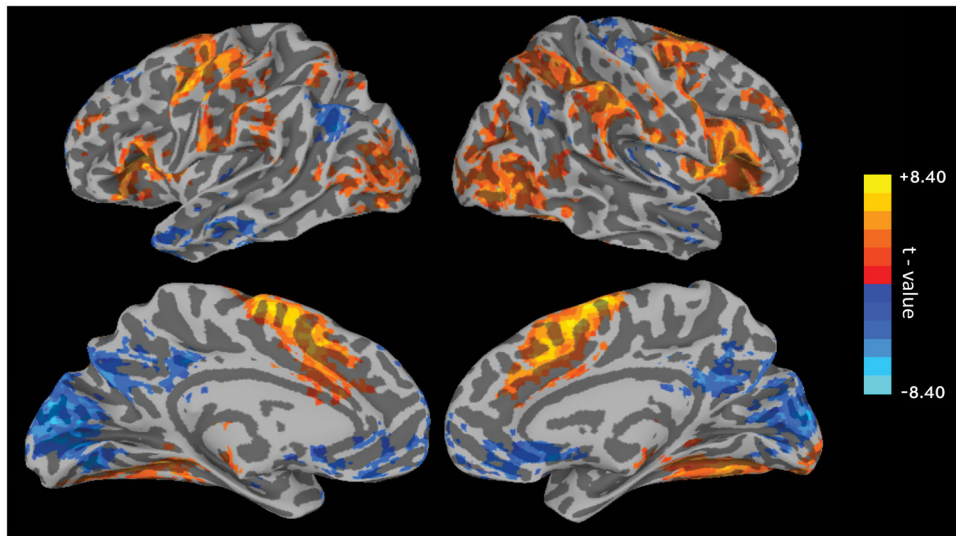
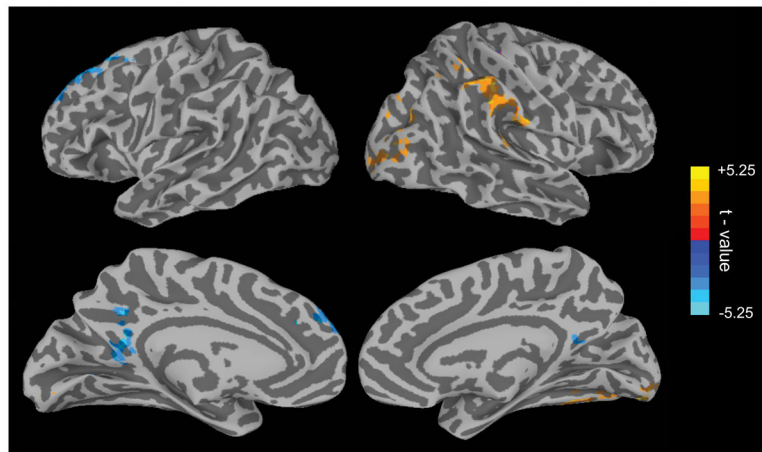


Figure 7. Variance Time Course Results

BOLD signal correlation with the reaction time variability time course. Positive (yellow) values show regions where BOLD activity increased as reaction time variability increased; negative (blue) values show regions where BOLD activity decreased as reaction time variability increased. The map shows T-statistics after correction for multiple comparisons (corrected $p < 0.05$; nominal $p < 0.01$, cluster size > 81 voxels).

**Figure 8. Variance Time Course and Performance Results**

Results of analysis correlating individual VTC-BOLD signal correlation strength with overall performance (d') on the task across participants. The map shows T-statistics after correction for multiple comparisons (corrected $p < 0.05$; nominal $p < 0.01$, cluster size > 81 voxels).

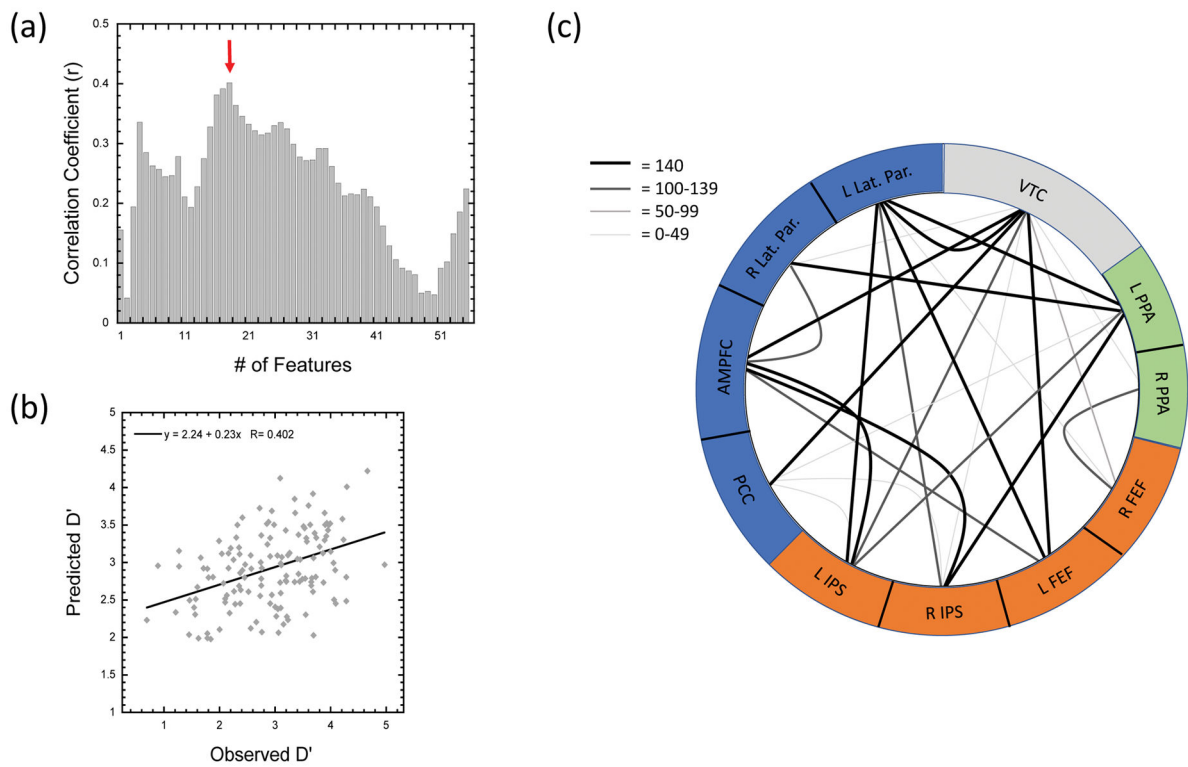


Figure 9. Predicting Individual Differences in Performance

Results of the leave-one-subject-out multiple linear regression analysis. (a) Pearson's r correlation coefficient between the measured and predicted d' score as a function of the number of features (connections) used in the model. (b) Scatterplot showing the relationship between the measured and predicted d' scores for all 140 participants in the 18-feature regression model. (c). Figure showing the 25 connections selected at least once between the 10 nodes of the PPA, DMN, and DAN ROIs, and the VTC in the 18-feature regression model. The color and weight of the connection show the number of folds that each of the connections was selected as a feature in the prediction model (max = 140 folds).

Table 1

Group means on the behavioral performance measures in the present study. For comparison, the mean values from the 1st run of the 16 participants from Esterman et al. (2013) are shown and compared to the means of the present sample. One-sample t-tests on the slope parameters tested if the slopes differed from zero (i.e., did performance change over time).

Parameter	Mean \pm 95% CI	1-Sample t-test ($x_0 = 0$)	Esterman (2013) 1 st run (95% CI)	Mann-Whitney Across samples
Mean reaction time (RT)	0.764 \pm 0.012		0.780 \pm 0.037	$U = 991, p = 0.451$
Reaction time variability (CV)	0.197 \pm 0.007		0.187 \pm 0.024	$U = 910, p = 0.220$
Commission Error Rate (CE)	0.212 \pm 0.023		0.271 \pm 0.065	$U = 809, p = 0.069$
Omission Error Rate (OE)	0.050 \pm 0.013		0.043 \pm 0.026	$U = 1109, p = 0.949$
D'	2.926 \pm 0.141		2.695 \pm 0.485	$U = 951, p = 0.324$
Criterion	0.573 \pm 0.071		0.684 \pm 0.117	$U = 1002, p = 0.491$
Slope - RT	-0.0015 \pm .00015	$t(139) = -2.016, p = 0.046$	0.0049 \pm 0.0043	$U = 694, p = 0.013$
Slope - CV	0.0046 \pm 0.0013	$t(139) = 6.938, p < 0.0001$	0.0042 \pm 0.0030	$U = 1065, p = 0.748$
Slope - CE	0.0139 \pm 0.0051	$t(139) = 5.383, p < 0.0001$	0.0046 \pm 0.0113	$U = 980, p = 0.413$
Slope - OE	0.0046 \pm 0.0024	$t(139) = 3.792, p < 0.0001$	0.0045 \pm 0.0050	$U = 1101, p = 0.909$
Slope - D'	-0.0898 \pm 0.0237	$t(139) = -7.426, p < 0.0001$	-0.0661 \pm 0.0439	$U = 1025, p = 0.579$
Slope - Criterion	0.0048 \pm 0.0115	$t(139) = 0.815, p = 0.416$	-0.0199 \pm 0.0313	$U = 976, p = 0.400$

Table 2

Statistical results of correlations across the four primary performance measures, mean reaction time, reaction time variability (CV), d' , and criterion shown in the scatterplots of Figure 2. The bottom row compares the difference in the strength of the correlation in the two rows above.

	Accuracy Measures	
	D'	Criterion
Mean Reaction Time	$r = -0.490, p < 0.001$	$r = -0.705, p < 0.001$
Coefficient of Variation	$r = -0.749, p < 0.001$	$r = -0.288, p = 0.001$
Difference of Correlation	$z = -3.943, p < 0.0001$	$z = -5.566, p < 0.0001$

Table 3

Results of the confirmatory factor analysis testing the proposed two factor model: ability (reaction time variability and d') and strategy (mean reaction time and criterion). We considered a ratio of χ^2 to degrees of freedom = 3, SRMR and RMSEA values = .06, and CFI and TLI values > .95 to be indicators of good model fit (Hu & Bentler, 1999; Schreiber et al., 2006). All criteria indicated a good fit except for Root Mean Square Error of Approximation (indicated by * in the table).

Quality of Fit Measurement	Value
χ^2 /degrees of freedom	2.525*
Standardized Root-Mean-Square Residual (SRMR)	0.024*
Root Mean Square Error of Approximation (RMSEA)	0.104
Comparative Fit Index (CFI)	0.994*
Tucker-Lewis Index (TLI)	0.963*

Table 4

Pretrial Activity. The results below show the measured average activation level across the -1.6sec to -0.8sec window prior to target onset for the PPA, DAN, and DMN ROIs. The first four columns show data from the original Esterman et al. (2013) study. Mean activity level for Commission Errors (CE) and Correction Omissions (CO) are shown along with the mean and 95% confidence intervals for the CO-CE contrast. The last column shows the mean CO-CE contrast levels for the present study, showing that mean differences in the DAN ROI in the present study falls outside the 95% confidence interval from Esterman et al. (2013).

ROI	Esterman et al. (2013)				Current Study
	CE	CO	Difference (CO-CE)	Difference 95%CI	Difference (CO-CE)
PPA	-0.0016	0.0211	0.0227	0.0049 to 0.0405	0.0157
DAN	-0.0032	0.0129	0.0161	0.0043 to 0.0278	0.0041
DMN	0.0353	-0.0030	-0.0383	-0.0612 to -0.0154	-0.0228

Table 5

Summary table showing the number of connections across regions in the leave-one-subject-out linear regression analysis using VTC and network level coupling to predict individual participant performance (d') in the 18-Feature model. A total of 25 unique connections were selected at least once across the 140 folds of the 18-Feature model (see Figure 9c). The main table shows the distribution of the 25 connections across the 9 potential categories. The bottom row, labelled Node Total, shows the total number of times each of the three networks (PPA, DAN, DMN) or the VTC was selected as one of the two nodes in the 25 connections. Thus, the Node Total values are not equal to the sum of the column above and the sum across the Node Total row is 50 (2 nodes \times 25 node pairs).

	PPA	DAN	DMN	VTC
PPA	0	2	3	1
DAN		0	10	4
DMN			1	4
Node Total	6	16	19	9