Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought

Agnieszka Zuberer[1,2,3], Aaron Kucyi[4,11], Ayumu Yamashita[1,2], Charley M. Wu[5,6,7], Martin Walter[8,9,10,11], Eve Valera[12,13], and Michael Esterman[1,2,14]

[1]Department of Psychiatry, Boston University School of Medicine
[2]Boston Attention and Learning Laboratory, VA Boston Healthcare System
[3]Department of Psychiatry and Psychotherapy, Jena University Hospital
[4]Department of Psychology, Northeastern University
[5]Human and Machine Cognition Lab, University of Tübingen
[6]Department of Psychology, Harvard University
[7]Center for Adaptive Rationality, Max Planck Institute for Human Development
[8]Department of Psychiatry and Psychotherapy, University of Tübingen
[9]Clinical Affective Neuroimaging Laboratory, Otto-von-Guericke-University Magdeburg
[10]Department of Behavioral Neurology, Leibniz Institute for Neurobiology
[11]Department of Psychiatry, Harvard Medical School
[12]Department of Psychiatry, Massachusetts General Hospital
[13]National Center for PTSD, VA Boston Healthcare System

Sustained attention is a fundamental cognitive process that can be decoupled from distinct external events, and instead emerges from ongoing intrinsic large-scale network interdependencies fluctuating over seconds to minutes. Lapses of sustained attention are commonly associated with the subjective experience of mind wandering and task-unrelated thoughts. Little is known about how fluctuations in information processing underpin sustained attention, nor how mind wandering undermines this information processing. To overcome this, we used fMRI to investigate brain activity during subjects' performance (n=29) of a cognitive task that was optimized to detect and isolate continuous fluctuations in both sustained attention (via motor responses) and task-unrelated thought (via subjective reports). We then investigated sustained attention with respect to global attributes of communication throughout the functional architecture, i.e., by the segregation and integration of information processing across large scale-networks. Further, we determined how task-unrelated thoughts related to these global information processing markers of sustained attention. The results show that optimal states of sustained attention favor both enhanced segregation and reduced integration of information processing in several task-related large-scale cortical systems with concurrent reduced segregation and enhanced integration in the auditory and sensorimotor systems. Higher degree of mind wandering was associated with losses of the favored segregation and integration of specific subsystems in our sustained attention model. Taken together, we demonstrate that intrinsic ongoing neural fluctuations are characterized by two converging communication modes throughout the global functional architecture, which give rise to optimal and suboptimal attention states. We discuss how these results might potentially serve as neural markers for clinically abnormal attention.

Keywords: Integration, Segregation, Sustained attention, fMRI, spontaneous thought

Significance Statement

Most of our brain activity unfolds in an intrinsic manner, i.e., is unrelated to immediate external stimuli or tasks. Here we use a gradual continuous performance task to map this intrinsic brain activity to both fluctuations of sustained attention and mind wandering. We show that optimal sustained attention is associated with concurrent segregation and integration of information processing within many large-scale brain networks, while task-unrelated thought is related to

sub-optimal information processing in specific subsystems of this sustained attention network model. These findings provide a novel information processing framework for investigating the neural basis of sustained attention, by mapping attentional fluctuations to genuinely global features of intra-brain communication.

Introduction

Cognition arises from the interplay between short- and long-range communication throughout the brain, supported by a robust intrinsic network architecture typically measured during wakeful rest (Smith et al., 2009). These intrinsic network interdependencies fluctuate in an ongoing manner over seconds and minutes and are not locked to immediate sensory input (Coste et al., 2011; Sadaghiani & D'Esposito, 2015). One fundamental cognitive function that per definition is self-emergent and not bound to task structure is sustained attention , the ability to maintain attention over prolonged periods of time (Esterman & Rothlein, 2019; Fortenbaugh et al., 2015). Sustained attention varies widely from moment to moment, and its associated behavioral fluctuations are thought to be linked to fluctuations of this intrinsic network architecture (Esterman et al., 2013; Kucyi et al., 2018). Mind wandering, often referred to as stimulus-independent and task-unrelated thought, is a common correlate of sustained attention and accompanies attentional lapses (Smallwood & Schooler, 2006), as reflected in more variable behavior or erroneous responses during sustained attention tasks (Bastian & Sackur, 2013), a go/no-go task (Stawarczyk et al., 2011) and an executive-control task (McVay & Kane, 2009). However, it is unresolved how the neural signature of objective measures of sustained attention is affected by mind wandering (Kucyi et al., 2016).

Sustained attention is not reliant on a limited set of specific brain regions but instead rather relies on the coordination of many large-scale networks (D. Godwin et al., 2015; Posner & Rothbart, 2007; Sadaghiani et al., 2015). With regard to mind wandering, on the one hand it is frequently linked to the default mode network: state mind wandering has been linked to the default mode network (Andrews-Hanna et al., 2010; Christoff et al., 2009; Mittner et al., 2014; Stawarczyk et al., 2011), while trait mind wandering is related to connectivity between the DMN and fronto-parietal control network as well as to connectivity within the DMN (C. A. Godwin et al., 2017; Golchert et al., 2017; Kucyi & Davis, 2014; OCallaghan et al., 2019). However, there is work arguing that task unrelated thought is related to activation in regions adjacent to unimodal sensorimotor cortex (Sormaz et al., 2018). With regard to objective measures of sustained attention, several attempts have linked sustained attention to patterns of whole-brain node-to-node connectivity (Castellanos & Proal, 2012; Rosenberg et al., 2016). However, rather than reveal a genuinely global attribute of brain communication, these studies capitalize on machine learning and selection of predictive connections, agnostic to their implications for information processing.

Global attributes of intra-brain communication have been less commonly investigated with regard to sustained attention. These attributes are generally mapped with graph theoretical approaches, e.g., by means of global efficiency (Stanley et al., 2015) or graph spectral entropy (Sato et al., 2013). An alternative and prominent approach is to map information processing as the degree of connectivity within tight-knit communities, often measured with the within-module degree z-score (Fortunato, 2010; Newman, 2006) and the diversity of connectivity across those communities, as sometimes measured by the participation coefficient. Thus, these measures are complementary, and a potentially ideal approach to map both global and local aspects of the intrinsic functional architecture. While a variety of integration measures have been used to investigate the network structure during cognitive tasks, the participation coefficient has become increasingly popular. However, few studies have investigated neural information processing with respect to both integration and segregation. (Cohen & D'Esposito, 2016; Shine et al., 2016). While in the majority of cognitive tasks, higher integration levels favor better performance (Shine & Poldrack, 2018), vigilance and motor learning tasks have favored enhanced segregation (Shine et al., 2016 and Bassett et al., 2015, respectively). In fact, Shine and Poldrack (2018) also stressed that some tasks may rely on higher segregation, such as sustaining attention, and that the role of integration/segregation may vary across different aspects of attention, given its multiple facets (directed vs diffuse; endogenous vs exogenous; local vs global; overt vs covert; visual vs corporeal; see Shine & Poldrack, 2018). Critically, research relating integration and segregation to cognitive functions that are not bound to the task structure (i.e., vigilance, sustained attention or mind wandering) is quite limited. One study found attentional lapses to be related to reduced segregation of the default mode network (DMN) and visual network, which was interpreted as favored encapsulation from other networks (Sadaghiani et al., 2015).

To conceptually advance our understanding of how sustained attention is linked to these global attributes of intra-brain communication, and thus modes of information processing, we utilize a unique gradual onset continuous performance task developed to map ongoing intrinsic fluctuations of sustained attention and mind wandering during fMRI (Esterman et al., 2013; Esterman et al., 2014; Fortenbaugh et al.,

Correspondence concerning this article should be addressed to Agnieszka Zuberer, Department of Psychiatry, Boston University School of Medicine, Boston, MA 0213. E-mail: azuberer@gmail.com

2015; Fortenbaugh et al., 2018; Kucyi et al., 2016). Subjects had to constantly respond to frequent targets and withhold response for rare targets while stimuli slowly transitioned. This way, stimulus-evoked effects were kept minimal and created a constant (background) state of readiness.

This design enabled us to investigate how fluctuations of sustained attention relate to fluctuations of two brain information processing modes: *segregation*, defined as the degree of within-community functional connectivity, and *integration*, defined as the diversity of cross-community functional connectivity. Our findings suggest that mapping these information processing modes via large-scale intrinsic network interdependencies can provide novel insights into how they underpin both fluctuations of sustained attention and mind wandering.

Material and Methods

Participants

Twenty-nine healthy, right-handed adults (13 males, 16 females; mean age = 26.7; SD= 3.9) provided written informed consent for procedures, which was approved by the Partners Human Research Institutional Review Board. The gradCPT was presented with nine self-paced thought-probes in each of four fMRI runs, using a 3T Siemens CONNECTOM scanner with 64-channel head coil. These data were published in a previous manuscript (Kucyi et al., 2016) using completely orthogonal set of preprocessing, analyses and research aims. Additional study details can be found in (Kucyi et al., 2016).

Preprocessing

We performed preprocessing of the fMRI data using FMRIPREP version 1.3.0 (Esteban et al., 2019). Preprocessing steps included realignment, co-registration, segmentation of T1-weighted structural images, normalization to Montreal Neurological Institute (MNI) space. Many internal operations of FMRIPREP use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details about the pipeline see https://fmriprep.readthedocs.io/en/latest/workflows.html. The BOLD signal time courses were extracted with spatial smoothing using an isotropic Gaussian kernel of 5 mm full-width at half-maximum from 264 ROIs. To remove several sources of spurious variance, we used linear regression with 9 regression parameters, including six motion parameters, average signals over the whole brain, white matter, and cerebrospinal fluid. We finally applied a high-pass (0.01 Hz cutoff) temporal filter to remove intrinsic scanner-related low frequency signal drift. The software nilearn/scikit-learn in Python (Abraham et al., 2014) was used for denoising. For the analysis of connectivity matrices, region of interests (ROIs) were delineated according to a 264-node gray matter atlas (Power et al., 2011). The 264 ROIs were defined as 4-mm spheres around the center coordinates that were determined in the previous studies (Power et al., 2011).

Task paradigm

We utilized a well-validated sustained attention task (Esterman et al., 2013; Fortenbaugh et al., 2015; Fortenbaugh et al., 2018; Kucyi et al., 2016) with a gradually changing background state that maps attentional readiness for rare targets by monitoring both steady motor responses and subjective reports via intermittent thought probes (~ every 44-60 seconds): as shown in Figure 1, gradually transforming city scenes (frequent, ~90%) and mountain scenes (rare; ~10%) were presented across four runs (~9 min each), while subjects were instructed to respond to frequent city scenes and withhold response to rare mountain scenes. Transient task-evoked influences were kept minimal due to varying long intervals between targets while creating a continuous background state with slow gradual transitions (vs. abrupt onsets) between scenes. The constant motor response enabled us to map ongoing attentional fluctuations via behavior variability, previously shown to be related to optimal (in the zone) and suboptimal (out of the zone) states of sustained attention (Esterman et al., 2013; Esterman et al., 2014). Further, we used an experience sampling approach with intermittent thought probes assessing subjective states of mind wandering, previously shown to uniquely associate with brain activity above and beyond task performance (Kucyi et al., 2016). Each of the four runs was interrupted 9 times every 44-60s by the thought probes.

Graph Construction

We use a graph analytic approach to describe information processing throughout the functional connectome. A graph $\mathcal{G}$ is defined by set of nodes $\mathcal{V}$, connected by edges $\mathcal{E}$. Nodes represent the brain regions defined by a brain parcellation and the edges represent the connection between those nodes.

Each edge e_{ij} defines the functional connectivity between nodes i and j, based on the Pearson correlation of the BOLD time series. The Pearson correlation coefficients are Fisher z-transformed, and all subsequent graph analyses were performed on the resulting 36 signed, unthresholded adjacency matrices A_T for each individual subject, after removing negative edges. This is calculated individually for each subject and at each time T defined by the 36 pre-thought probe blocks. These probes are of variable length (44-60s), thus our analyses use the 40s interval before each probe, resulting in 1044 pre-thought probe blocks for all subjects (n=29). Nodes are defined using the parcellations of the Power atlas (Power et al., 2011), which describes 264 distinct brain regions that have high homogeneity and do not share physical boundaries. These specific qualities avoid overestimating the local connectivities between brain regions. The Power atlas

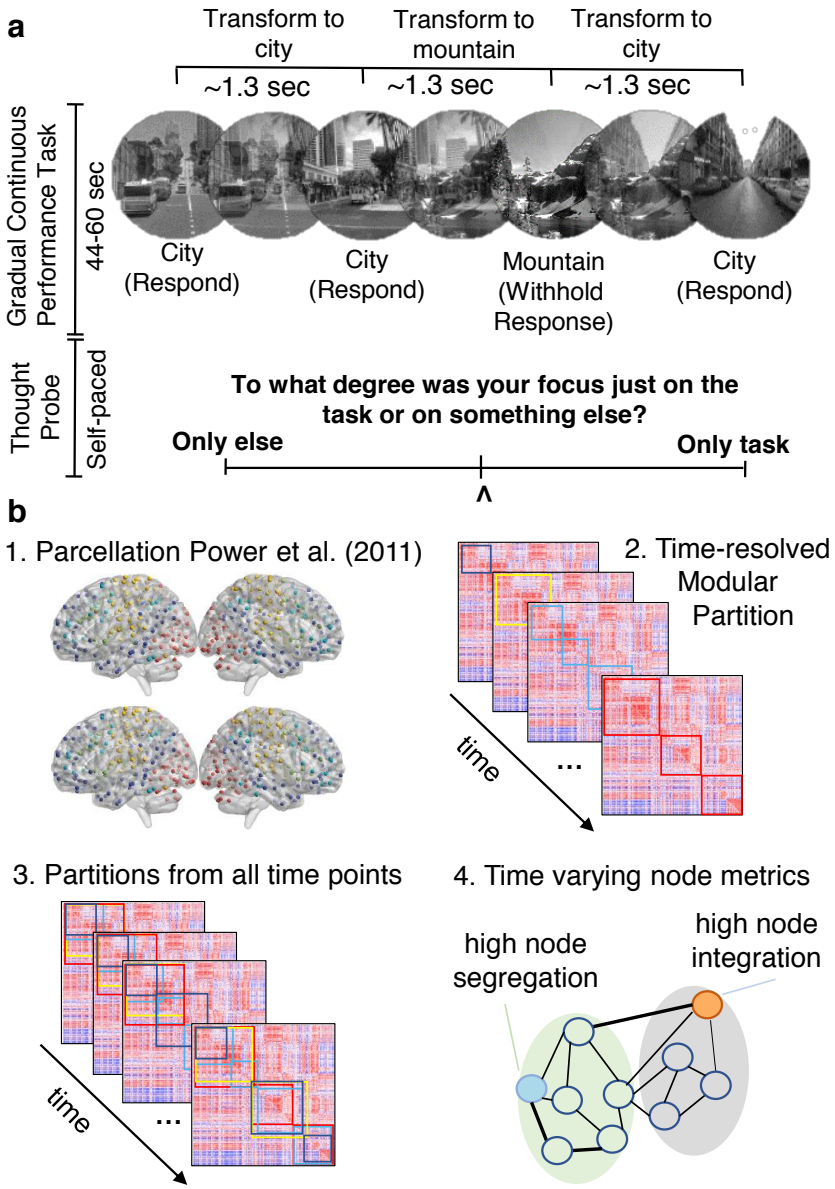

Figure 1

Schematic illustration of task design and time-resolved graph construction. **(a)** Task paradigm. Subjects were instructed to respond to frequent gradually transforming city scenes. At unpredictable long intervals, rare targets (mountain scenes) required response inhibition. Every 44–60s, a self-paced thought-probe (bottom) is displayed instructing subjects to evaluate the degree of task focus just before appearance of the thought-probe on a continuous scale (100 for "only task" and 0 for "only else"). **(b)** Node parcellations were based on an atlas by Power et al. (2011) comprising cortical, subcortical and cerebellar regions (264 regions in total). Community detection was performed for every pre-thought probe segment (40s) across all 36 task blocks separately. Node metrics were derived based on all possible partitions that were constructed for each block separately.

is one of few atlases that is defined based both on functional connectivity and studies of task activations, which is especially important for our present analyses. Further, this atlas includes cortical, subcortical, and cerebellar regions. It accurately assigns nodes into communities observed with other approaches (e.g., at the voxel level), and these communities have been widely used (Cole et al., 2013; Gu et al., 2015; Power et al., 2011; Power et al., 2013).

Time-resolved community structure assignment

We performed a community detection based on the Louvain algorithm (Blondel et al., 2008) implemented in the Brain Connectivity Toolbox (Rubinov et al., 2009). A modularity statistic Q is iteratively maximized for different

community assignments until Q reaches its possible maximum. In this respect, the modularity estimate Q represents a measure for the extent to which a graph can be subdivided into community structures that display stronger within-community connectivity than cross-community connectivity.

$$Q_t = \frac{1}{d^+} \sum_{ij} \left(w_{ij}^+ - e_{ij}^+\right) \delta_{M_i M_j} - \frac{1}{d^+ + d^-} \sum_{ij} \left(w_{ij}^- - e_{ij}^-\right) \delta_{M_i M_j} \tag{1}$$

Equation 1 shows the Louvain modularity algorithm, where d is the total weight of the network (sum of all connections), w_{ij} is the weighted and signed connection between nodes i and j, e_{ij} is the strength of a connection divided by the total weight of the network, and $\delta_{M_i M_j}$ is set to 1 when regions are in the same community and 0 when they are not in the same community. The super-scripts + and − denote all positive and negative connections, respectively. The Louvain algorithm was applied 100 times on the same adjacency matrix and across 60 different spatial resolution levels γ, which we implemented using $0.1 < \gamma \geq 6$ in steps sizes of 0.1, with the final partition decided based on the consensus partition algorithm, with the highest agreement by the mutual information criterion.

Integration

After assigning nodes to communities as described in the above section, we calculated the participation coefficient B_T (Guimera & Amaral, 2005) for each node i as a measure of the integration of information processing:

$$B_{iT} = 1 - \sum_{s=1}^{N_M} \left(\frac{K_{i_{sT}}}{K_{iT}}\right)^2. \tag{2}$$

$K_{i_{sT}}$ is the degree of the positive connections of region i to regions within its community s at time T, K_{iT} is the degree of all positive connections of region i at time T, and N_M is the total number of communities.

Intuitively, the participation coefficient B_T measures the distribution of edges of a node among the communities of a graph. A node's participation coefficient maximally approaches 1 if the sum of edge weights to each community are equally distributed.

Segregation

To model the segregation of information processing, we use the strength of within module connectivity, as defined by the module-degree z-score W_T (Guimera and Amaral, 2005), where for each node i:

$$W_{iT} = \frac{K_{iT} - \bar{K}_{s_{iT}}}{\sigma(K_{s_{iT}})}, \tag{3}$$

where $\bar{K}_{iT}$ is the degree of connections of region i to other regions in its module, s_i, at time T, $K_{S_{iT}}$ is the average of k across all nodes in s_i and $\sigma(K_{S_{iT}})$ is the standard deviation of k in s_i at time T. Intuitively, the within module degree z-score W_T measures the strength of connections a node i has to other nodes within its community relative to other nodes in their community.

Time-varying partitions

Thompson et al. (2020) showed that the mapping of nodal properties (e.g., segregation or integration) can lead to misleading results, when not taking into account the temporal fluctuations of communities. To address this issue, we computed our nodal properties as the mean across all possible community partitions across all time points (i.e., 36 pre-thought probe blocks, each lasting 40s/37 volumes).

Behavioral predictors

Optimality of sustained attention was operationalized as reaction time variability, derived from the latency of correct responses to the frequent city scenes. Error responses (i.e., to rare mountain scenes) were excluded from this analysis. Reaction time variability is a widely established marker of sustained attention, and has also been used in gradual continuous performance tasks (Esterman et al., 2013; Fortenbaugh et al., 2015; Fortenbaugh et al., 2018; Kucyi et al., 2016). The degree of mind wandering was operationalized by self-reports during task-intermittent thought probes, where subjects had to rate the degree to which their focus was on task or on something else on a scale between 1-100 (see section *Task paradigm*). Thought probes of this kind have been widely used to map mind wandering (Christoff et al., 2009; Kucyi et al., 2016; Stawarczyk et al., 2011).

Statistical analysis

Linear mixed-effects (LME) regression (Baayen et al., 2008) models were used to predict measures of global brain-connectivity based on behavioral markers of sustained attention. Specifically, we fit two separate linear mixed effects models (model I and II; see Table A2), predicting one of two dimensions of information processing: integration (B_T; model I) and segregation (W_T; model II). The model's dependent variables (integration and segregation) are based on all 36 pre-thought probe blocks, derived on a subject level, resulting in 1044 pre-thought probe blocks in total. By using subject assignment as a random intercept, inter-individual differences in integration and segregation were taken into account. We followed a step-wise approach, including a predictor if it significantly improved the model fit ($p < .05$), using an ANOVA on the log-likelihood ratio of the two models. For interactions, the main effects were kept in the model following the principle of marginality. Statistical analysis was performed using the `lme4` package (Bates et al., 2014) in **R**. As possible predictors, we used the putative assignment of

nodes to higher-order systems (i.e., canonical networks comprised of multiple nodes) provided by the Power atlas (Power et al., 2011). All behavioral variables were centered within subjects and the dependent variables (segregation and integration, respectively) were referenced to the grand mean, i.e. the mean of segregation or integration across all 264 nodes.

As a behavior predictor, we used the standard deviation of reaction times (RT-SD) for each pre-though probe block. This measures behavioral variability, which is a widely established marker of sustained attention (Esterman et al., 2013; Fortenbaugh et al., 2015; Fortenbaugh et al., 2018; Kucyi et al., 2016). We refer to these two models as sustained attention models. In subsequent analyses, we investigated whether the two sustained attention models were improved by the inclusion of degree of task-unrelated thought (mind wandering) as a predictor, since the degree of mind wandering is known to be a contributor to sustained attention (Barkley, 1997; Kucyi et al., 2016). We further investigated head motion as a possible confound (see section *head motion*). The final models predicting integration or segregation respectively included a 2-way interaction between behavioral variability (RT-SD) and network assignment (Table A1). Further, degree of mind wandering significantly interacted with behavioral variability and network assignment resulting in a 3-way interaction between, RT-SD, degree of mind wandering and network assignment (Table A2).

Confidence intervals were bootstrapped with 1000 iterations and p-values were computed via Wald-statistics approximation (treating t as Wald z). While we report p-values, note that significance testing and the interpretability of resulting p-values is highly debated in mixed-effects modeling (for discussion see Baayen et al. (2008)). Previous work has suggested that correction for multiple comparisons is not mandatory within mixed-effects modeling (Gelman et al., 2012), however we report p-values with and without correction for multiple comparisons. Corrections for multiple testing were performed with the false discovery rate (FDR) procedure for 11 hypotheses (11 networks) (Benjamini & Yekutieli, 2001).

Rich/diverse hubs of sustained attention

Rich hubs of optimal attention refer to nodes that exhibit high within-network connectivity (segregation) with optimal attention, and diverse hubs of optimal attention refer to nodes exhibiting high between-network connectivity (integration) with optimal attention. Diverse and rich hubs of low sustained attention vs. high sustained attention were based on a cutoff of our estimates of the interaction term between node assignment and reaction time variability (positive and negative association) of a linear mixed effects model predicting the node integration B_T or node segregation W_T. We determined knee points as cut-offs where the sorted estimates showed a steep increase and steep decrease. The knee point function in Matlab operates by walking along the curve by one bisection point at a time. It fits two lines, one to all the points to left of the bisection point and one to all the points to the right of the bisection point. Finally, the knee point is judged to be at a bisection point which minimizes the sum of errors for the two fits.

Results

We utilized a well-validated sustained attention task (Esterman et al., 2013; Fortenbaugh et al., 2015; Fortenbaugh et al., 2018; Kucyi et al., 2016) creating a background state to map both sustained attention and task-unrelated thoughts to two global information processing modes- segregation and integration- using graph analyitic tools.

Task performance

During gradCPT, fluctuating levels of reaction time (RT) variability are a marker of attentional state (Esterman et al., 2013; Esterman et al., 2014; Fortenbaugh et al., 2018). As is commonly observed, subjects with higher RT variability had higher rates of attentional lapses, or errors to rare mountain targets (r = .69, p < .001). As observed previously, subjects with higher mean off-task rating across all thought-probes had higher reaction time (RT) variability (r = 0.58, p < .001) and attention lapse (error) rate (r = .36, p = .058), but not differential mean RT (r = .11, p = 0.58; see Fig. A1). Of particular relevance to the current study, on a within-subject level, the degree of mind wandering was positively related to behavioral variability (R = .16) and this effect was significantly different from zero based on a one sample t-test (T_{28}, p = .003; see Fig. A3), as shown previously in this data set (Kucyi et al., 2016). This suggests that within individuals, behavioral variability was modestly coupled with degree of mind wandering. Together these findings suggest that RT variability is an objective marker of attentional state, and further that mind wandering contributes to suboptimal sustained attention.

Graph Analysis

In order to determine how fluctuations in sustained attention relate to alterations of the intrinsic functional architecture, we isolated time series segments of 40 seconds preceding each thought probe (from here on called block), resulting in a total of 36 blocks per subject (4 runs with 9 blocks each). For each block and subject separately, we then used graph theoretical analysis to identify the community structure of the brain connectome (Rubinov & Sporns, 2010). Next, we determined a) the diversity of the distribution of every brain region's connections, or integration between these communities quantified by the participation coefficient (B_T) and b) the degree of connections within each of the communities, or segregation quantified by the modularity degree

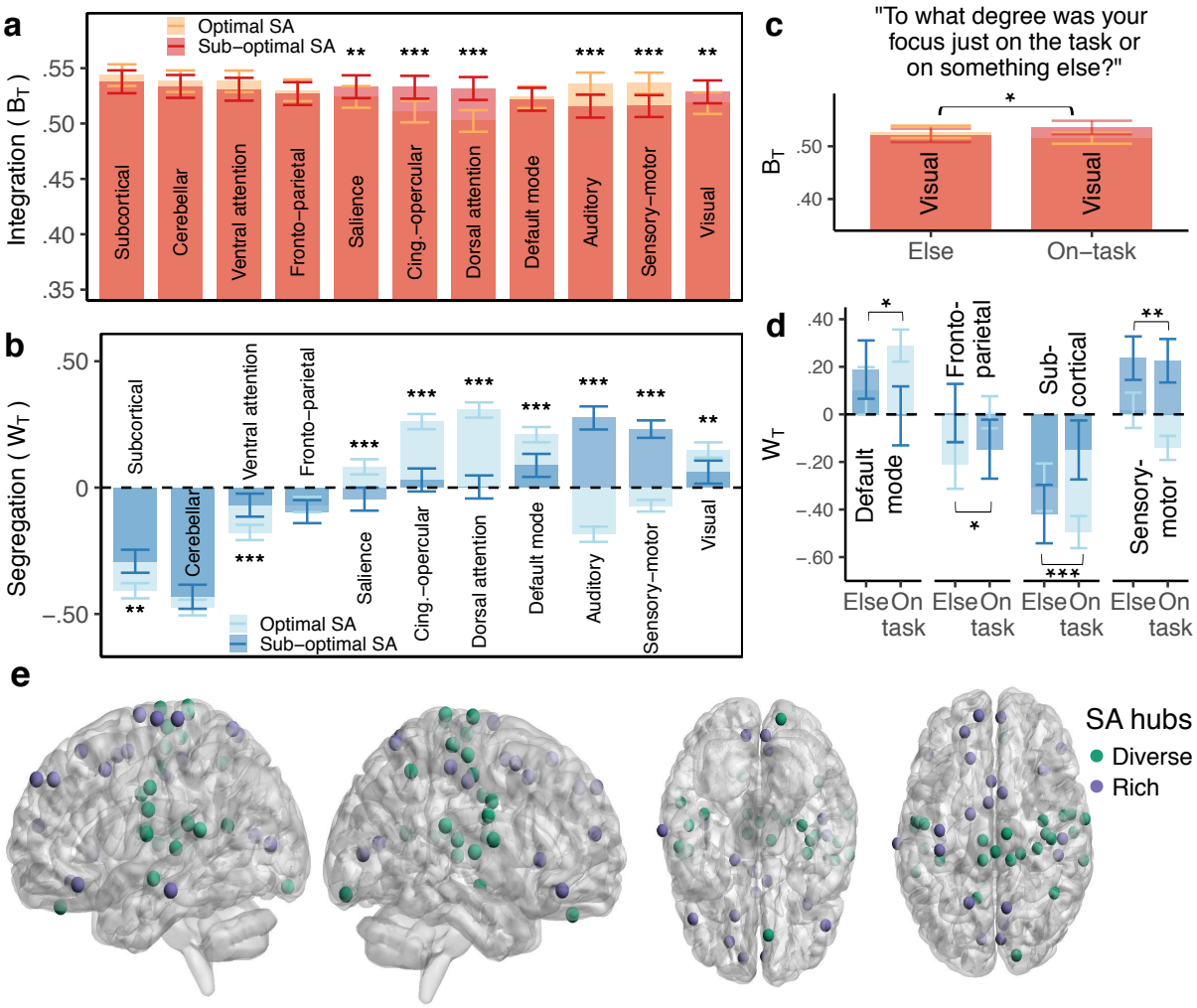

Figure 2

Sustained attention network (SAN) model based on two modes of information processing across the whole-brain connectome, the diversity of cross-community connectivity (B_T) and degree of within-community connectivity (W_T). **(a)** and **(b)** display the interaction effect of RT variability (dichotomized into optimal and sub-optimal sustained attention, corresponding to low and high variability, respectively) and network assignment for a linear mixed effects model predicting B_T and W_T. The dichotomization of RT variability (performed using a median split) was only for visualization purposes, whereas the original models used a continuous variable. **(c)** and **(d)** display the interaction effect of RT variability, degree of mind wandering, and network assignment for a linear mixed effects model predicting B_T and W_T, respectively. **(e)** Brain maps showing nodes that exhibit over-proportionally enhanced integration and reduced segregation (green; here called diverse hubs of sustained attention) and nodes that exhibit over-proportionally enhanced segregation and reduced integration (purple; here called rich hubs of sustained attention). Significance levels: $*p < .05$, $**p < .01$, $***p < .001$.

(W_T; Guimera & Amaral, 2005); see Fig. 1 and Materials and Methods). Integration and segregation were moderately correlated within subjects (mean $R = -0.43$) and this effect was significantly different from zero (one sample t-test; T_{28}, $p < .001$; Fig A2). There was only a weak between-subjects correlation between integration and segregation (R=-0.06) respectively. The results suggest that integration and segregation convey non-redundant information, although they are related to each other.

Sustained Attention network model

First, we investigated if the segregated and integrated information processing modes within large-scale networks across the whole-brain connectome varied with fluctuations of sustained attention. There was a significant interaction between network assignment and reaction time variability for both models predicting the integration (B_T) and segregation (W_T) of information processing, respectively (Table A1 and Figure 2).

All networks that showed a significant interaction with at-

tention in the model predicting integration (B_T; Fig. 2a and Table A1) respectively, showed the opposing directionality in the model predicting segregation (W_T; Fig. 2b). Specifically, with reduced RT variability (optimal attention), networks exhibited enhanced integration and reduced segregation in the auditory ($\beta_{B_T} = -.09$; 95% CI: $[-.13, -.06]$, $p < .001$, $p_{FDR} < .001$; $\beta_{W_T} = .26$; 95% CI: $[.23, .29]$, $p < .001$, $p_{FDR} < .001$) and sensorimotor networks ($\beta_{B_T} = -.09$; 95% CI: $[-.12, -.07]$, $p < .001$, $p_{FDR} < .001$; $\beta_{W_T} = .18$; 95% CI: $[.16, .20]$, $p < .001$,$p_{FDR} < .001$; see Figure 2a).

Conversely, optimal attention was reflected in concurrently reduced integration and enhanced segregation in the salience ($\beta_{B_T} = .06$; 95% CI: $[.02, .10]$, $p = .001$;, $p_{FDR} = .002$ $\beta_{W_T} = -.07$, 95% CI: $[-.10, -.04]$, $p < .001$, $p_{FDR} < .001$), cingulo-opercular ($\beta_{B_T} = .13$, 95% CI: $[.09, .16]$, $p < .001$;, $p_{FDR} < .001$ $\beta_{W_T} = -.13$, 95% CI: $[-.16, -.10]$, $p < .001$, $p_{FDR} < .001$), dorsal attention ($\beta_{B_T} = .17$; 95% CI: $[.13, .20]$, $p < .001$;, $p_{FDR} < .001$ $\beta_{W_T} = -.17$, 95% CI: $[-.20 - .14]$, $p < .001$, $p_{FDR} < .001$), and visual networks ($\beta_{B_T} = .07$; 95% CI: $[.03, .10]$, $p < .001$, $p_{FDR} = .001$; $\beta_{W_T} = -.05$, 95% CI: $[-.08, -.02]$, $p = .002$, $p_{FDR} = .002$; see Figure 2b).

Additional networks showed an interaction between segregation of information processing and attentional state. Specifically, optimal attention predicted lower segregation in the subcortical and ventral attention networks (subcortical: $\beta_{W_T} = .07$, 95% CI: $[.04, .10]$, $p < .001$, $p_{FDR} < .001$; ventral attention: $\beta_{W_T} = .06$, 95% CI: $[.03, .09]$, $p < .001$, $p_{FDR} < .001$).

Overall, these findings indicate that optimal sustained attention arises from reduced network cross-talk (integration) and greater within-network communication (segregation) in task-relevant networks, including salience, cingulo-opercular, dorsal attention, and visual. In contrast, optimal attention predicted greater network cross-talk (integration) and reduced within-network communication (segregation) in auditory and sensorimotor networks. Additionally, the results indicate that both communication modes are inversely related to each other, such that networks exhibiting higher/lower integration inversely show lower/higher segregation varying with sustained attention. While both information processing modes converge with respect to many putative systems, the segregated mode of information processing extends to additional systems, not apparent for the integrated mode of information processing, including the subcortical and ventral attention networks.

Contribution of mind wandering on SAN model

Now that we established our global SAN model with an objective measure of sustained attention, we investigated how this neural signature of sustained attention is further explained by self-reports of mind wandering, one of many contributors to sustained attention (Esterman & Rothlein, 2019). In line with previous research, we used intermittent thought probes asking subjects to rate on a continuous scale between 0 and 100 the degree of their on-task focus ("To what degree was your focus on the task or on something else?"). The degree of mind wandering significantly interacted with reaction time variability and network assignment of nodes, both within our integration (B_T) and segregation (W_T) models. Specifically, in our integration model, only the visual system interacted significantly with RT variability and mind wandering, in that the favored reductions of integration levels associated with optimal sustained attention were lost with higher mind wandering and preserved with lower mind wandering ($\beta_{B_T} = -.034$, 95% CI: $[-0.08, -.00]$, $p = .034$, $p_{FDR} = .377$; see Fig. 2c). In other words, when optimal performance accompanied greater mind wandering, reduced visual network integration was not observed.

In our segregation model the subcortical network lost its favored segregation mode (lower segregation with optimal attention) when degree of mind wandering was high ($\beta_{W_T} = -.05$, 95% CI: $[-0.08, -.02]$, $p < .001$, $p_{FDR}p = .001$; see Fig. 2d). Similarly, auditory and sensorimotor networks reduced their favored segregation mode when mind wandering was high (auditory: $\beta_{W_T} = -.03$, 95% CI: $[-.06, -.00]$, $p = .046$, $p_{FDR} = .072$; sensorimotor: $\beta_{W_T} = -.03$, 95% CI: $[-.05, -.01]$, $p = .005$, $p_{FDR} = .010$; see Fig. 2d). Additionally, the default mode network, which did not interact with optimal attention alone, showed a significant interaction with reaction time variability and mind wandering in that at low levels of mind wandering, optimal attention predicted higher segregation (akin to task-related networks), but at high levels of mind wandering this relationship was absent $\beta_{W_T} = .04$; 95% CI: $[.01, .07]$, $p = .01$, $p_{FDR} = .019$; Fig. 2d). Similarly, the fronto-parietal network showed a significant interaction with reaction time variability and mind wandering, in that at low levels of mind wandering, optimal attention predicted higher segregation (akin to task-related networks), but at high levels of mind wandering this pattern reversed ($\beta_{W_T} = .04$; 95% CI: $[.01, .07]$, $p = .015$, $p_{FDR} = .025$; see Fig. 2d). Thus, when optimal performance accompanied greater mind wandering, the favored pattern of segregation across a number of networks was muted or even reversed.

Taken together, these findings suggest that mind wandering drew on only a subset of systems within our SAN model, and higher mind wandering generally undermined their favored integration or segregation mode during optimal sustained attention. Thus, these information processing markers of optimal sustained attention were strongest when objective and subjective measures converged, albeit only in select networks. Similar patterns were observed using an alternative parcellation (Schaefer et al., 2018), particularly with regard to segregation (see Supplementary Figure A4).

Head motion

Head motion can be a significant confound in graph analyses (Power et al., 2014; Siegel et al., 2017). In the present study, frame-wise displacement was relatively low ($0.1mm \pm .03mm$). To investigate, if our results were possibly confounded with head motion, we constructed a mixed-effects model predicting integration or segregation separately with the inclusion of frame-wise displacement as a predictor for each individual block for each subject separately. We included subject assignment as a random intercept, thereby taking into account inter-subject variability. There was no significant effect for framewise displacement neither for segregation (p=0.5549) nor integration (p=0.5465). We further investigated if head motion would interact with fluctuations of attention. To test this, we investigated if frame-wise displacement for each single block would interact with reaction time variability. For the models predicting segregation or integration, there was no significant effect for framewise displacement (segregation: $p = .7667$, integration: $p = .0705$). We repeated the same analysis with degree of mind wandering and found no significant impact on the model fit (segregation: $p = .6011$, integration: $p = .5209$).

Rich and diverse hubs of sustained attention

Lastly, we sought to isolate nodes that were sensitive to fluctuations of sustained attention with respect to these two modes of information processing. To this aim, we constructed our models at the node level and used a ranked cut-off metric (see Material and Methods). Nodes that exhibited concurrently higher segregation and lower integration with optimal sustained attention (here called rich hubs of sustained attention) were predominantly located in 1) the left-hemispheric superior and inferior frontal lobe belonging to the default mode and cingulo-opercular network, 2) bilateral supplementary motor nodes within the cingulo-opercular and salience networks, 3) sensorimotor nodes within postcentral areas, and 4) the bilateral calcarine in the visual network (see Fig. 2e and table Table A3). Nodes that exhibited concurrently higher integration and lower segregation with optimal sustained attention (here called diverse hubs of sustained attention) were predominantly located in bilateral para/postcentral regions within the sensorimotor network, and within the rolandic operculum and temporal regions belonging to the auditory network (see Fig. 2e and Table A4). Overall, these findings are consistent with our network SAN model, that optimal sustained attention arises from reduced network cross-talk (integration) and greater within-network communication (segregation) in rich hubs predominantly in task-relevant networks. Additionally, optimal sustained attention significantly predicted increased network cross-talk (integration) and reduced within-network communication (segregation) in diverse hubs predominantly in auditory and sensorimotor networks.

Discussion

In this study we investigated how sustained attention and mind wandering - two cognitive entities that per definition can be self-emergent and not bound to task structure - relate to ongoing fluctuations of the intrinsic functional architecture. We formulate a sustained attention network (SAN) model derived from genuinely global attributes of brain communication, i.e., the segregation and integration of information processing throughout large-scale networks (Guimera & Amaral, 2005; Newman, 2006). Specifically, we show that optimal periods of sustained attention, defined objectively via task performance, were associated with increased segregation and decreased integration in several task-relevant regions alongside increased integration and decreased segregation in sensorimotor. Subjective rating of mind wandering was a moderator of several networks in this SAN model, such that greater mind wandering weakened the optimal information processing state.

Sustained attention network model

Our data show that intrinsic fluctuations of sustained attention were underpinned by both enhanced integration and reduced segregation of information processing within auditory and sensorimotor systems, with concurrent reduced integration and increased segregation within the salience, cingulo-opercular, dorsal attention and visual systems. These latter networks are the most consistently activated by rare task events during gradCPT (Esterman et al., 2013; Fortenbaugh et al., 2015; Fortenbaugh et al., 2013; Kucyi et al., 2017), and their activity fluctuates with attentional states. Thus, our results indicate an optimal information processing mode for these networks, that may underlie their successful deployment when in the zone as well in successful inhibition to rare no-go targets.

Until now, there has been no cohesive framework to describe how these two modes of brain-communication change as a function of cognition, as the majority of studies have focused on network segregation only (but see Bertolero et al., 2015; Crossley et al., 2013; Shine et al., 2016; Yeo et al., 2015). There is some theoretic (Dehaene et al., 1998) and empirical evidence suggesting that more complex tasks favor a more integrated network architecture as opposed to less cognitively demanding tasks which exhibit a more segregated network organization (Cohen & D'Esposito, 2016; Stevens et al., 2012; Yue et al., 2017). On the other hand, it has recently been suggested that complex tasks, when practiced, favor greater segregation in task-related networks (Finc et al., 2020), which may be akin to efficient performance during in the zone/low variability states (Esterman et al., 2014). In an auditory detection task designed to map intrinsic fluctuations of vigilance (Sadaghiani et al., 2015), attentional

lapses were related to reductions of segregation of the default mode network (DMN) and visual network, similar to our results. However, this is one of the first studies to consider how fluctuations in sustained attention and mind wandering, cognitive processes unbound to task structure, are linked to segregated and integrated organization of the connectome.

In our SAN model, the strongest task-related network effects were reflected in increased segregation and decreased integration of cingulo-opercular and dorsal attention networks with optimal sustained attention. There is a body of research suggesting that increased activity and functional connectivity within the cingulo-opercular network supports enhanced tonic alertness (Sadaghiani et al., 2015), the detection of rare targets, and error related task-set reconfigurations (Fortenbaugh et al., 2018) while increased activation and connectivity within the dorsal attention network is thought to reflect selective goal-directed attention (Corbetta & Shulman, 2002; M. D. Fox et al., 2006; Sadaghiani et al., 2015). As reflected in a number of fMRI studies, these attention processes play key roles in the current task since subjects had to maintain attention and alertness across longer periods of time, selectively deploy goal-directed attention to rare targets, and reconfigure task set in response to errors and suboptimal performance (Esterman et al., 2014; Esterman et al., 2017; Fortenbaugh et al., 2018). Specifically, fMRI during gradCPT has revealed that fluctuations between optimal and suboptimal performance (e.g., in versus out of the zone) are associated with activity fluctuations in cingulo-opercular, dorsal attention and default mode networks (Esterman et al., 2016; Esterman et al., 2013; Esterman et al., 2014; Fortenbaugh et al., 2018; Kucyi et al., 2018). Additionally, fMRI and electrophysiology suggest that anti-correlations between these task-related systems and the default mode network help support attention, especially with the dorsal attention network (Rothlein et al., 2018). However there is some evidence suggesting that greater activation within the default mode system supports optimal attention on the one hand, and mind wandering on the other (Esterman et al., 2013; Fortenbaugh et al., 2015; Kucyi et al., 2016; Kucyi et al., 2017; Weissman et al., 2006).

Of note, nodes that exhibited the strongest concurrent segregation and lowest integration during optimal states of attention (rich hubs of sustained attention) were located within the cingulo-opercular and default mode networks, in the frontal lobes. Previous research has suggested that nodes within the frontal lobe play a central role during challenging tasks that demand information manipulation and retention with integration levels of these nodes across large-scale networks reflecting "network flexibility" (Braun et al., 2015; Stuss, 2006). Recent work has shown that the degree of structural connections of the white matter fiber tracts from and to the superior frontal lobe play a central role in the brain's ease of state transitions (as measured by brain controllability; Jamalabadi et al., 2020), lending further credit to this region's role in enabling network flexibility. However, the preferred segregation of these nodes for optimal attention could reflect preferred stability, rather than flexibility, as well as suppression of possible inference of task-unrelated information during this relatively monotonous task.

On the other hand, the strongest network effects reflecting decreased segregation and increased integration with optimal sustained attention occurred in sensorimotor and auditory networks. This was also reflected in the diverse hubs of optimal sustained attention which predominantly fell within these networks. In contrast, these networks are not critical for optimal visual attention required to perform gradCPT. Overall, these systems may benefit from top-down control/suppression (via integration) and weaker within-network information processing, which could be an indication of task-irrelevant processing (Talsma et al., 2010; Zimmer et al., 2010). Further research is needed to elucidate the association between these global measures of brain communication and activation within, as well as functional connectivity between specific brain systems.

Association of SAN model with mind wandering

Specific networks within and outside our SAN model showed further interactions with mind wandering, a subjective measure of attention state, known to impact sustained performance. In particular, several networks showed reduced associations (integration or segregation) with objective markers of sustained attention when mind wandering was high. Specifically, lower integration levels in the visual system were associated with optimal attention, however, this was diminished with higher mind wandering, suggesting that visual system is sensitive to perturbations from mind wandering. This hypothesis finds support from work showing that higher fidelity of visual representations related to optimal sustained attention (Rothlein et al., 2018). Further, higher local centrality and network connectivity in visual regions were associated with task performance during a visual categorization task (Ekman et al., 2012). Additionally, higher perceptual processing disruptions, often referred to as perceptual decoupling have been reported to relate to non-deliberate mind wandering (Seli et al., 2015).

Similarly, the favored lower segregation mode with optimal attention in the subcortical network (lower and higher segregation with optimal attention respectively) was lost when mind wandering reports were high. Our finding that subcortical nodes lost their favored reductions of segregation levels during mind wandering could indeed be initial evidence for an overly automatic task processing mode, facilitating a non-deliberate train of thought which eventually gives rise to mind wandering. Previous research has highlighted the importance of subcortical regions in mind wandering and their potential role in implementing automatic

constraints (Christoff et al., 2016). Additionally, the default mode network, which did not interact with optimal attention alone, showed decreased segregation levels when mind wandering was high. The default mode activation is associated with both optimal (in the zone) and suboptimal (mind wandering) attention, cognitive measures that notably go in opposite direction (Esterman et al., 2013; Kucyi et al., 2016). While overall, DMN exhibited higher segregation with optimal attention, akin to task-related network effects, mind wandering eliminated this association. This indicates that when the DMN is occupied with mind wandering, its segregated state no longer reflects objective measures of sustained attention (variability, or in/out of zone performance). Thus, this information processing mode may help explain why DMN activity can be both optimal and suboptimal (Kucyi et al., 2016). Additionally, the fronto-parietal network, which was not related to sustained attention in our SAN model, exhibited reduced levels of segregation with optimal attention when mind wandering was high, and increased segregation with optimal attention when mind wandering was low. This network is commonly associated with mind wandering. A meta-analysis involving 24 functional neuroimaging studies of mind wandering reports a co-activation of both executive and default mode networks. The fronto-parietal network is thought to couple more with DMN during mind wandering to select task-unrelated thoughts (K. C. Fox et al., 2015). Interestingly, when mind wandering is low, the fronto-parietal network shows increased optimality with greater segregation, like other task-related networks. However, this reverses with high mind wandering, similar to the pattern observed in the DMN. In this respect, favorable anti-correlation between both systems during low mind wandering periods could potentially be reflected in our "anti-segregation" pattern exhibited by both systems in our model of optimal sustained attention.

Limitations

It is unresolved to which degree the present results generalize to other tasks or scenarios. On the one hand, the lack of abrupt onsets in the gradCPT, versus more typical abrupt paradigms, may have helped to isolate these intrinsic fluctuations, by reducing exogenous onset cues to distinct trials, a major advantage of this task. In this respect, the present task is optimized to assess intrinsic, "self-emergent" fluctuations in attention. Nonetheless, errors and idiosyncratic task/stimulus properties, or fluctuations in (intrinsic) motivation could influence attention and we note that while the task attempts to minimize these factors, we cannot fully rule them out. The present gradCPT task is interrupted every ~45s instructing subjects to rate their current state of mind. In this respect, subjects might be constantly in a monitoring mode throughout the task, which may have undermined the examination of intrinsic fluctuations of attention. Generalization across alternative sustained attention paradigms is needed. Further, this work faced the problem of limited prior research examining attention fluctuations and fMRI measures of segregation and integration, thus formulating a priori hypotheses was difficult. Additional research is needed to show how the present findings are related to more traditional activation and connectome-based models of sustained attention, as well as how these approaches can complement each other. This SAN model should be validated in other tasks and subjects, as well as tested as predictors of individual and clinical differences in sustained attention. The results should be replicated with larger sample size. We chose the Power atlas (Power et al., 2011) based on a priori citeria (see Material and Methods). However, similar but not identical results were obtained with a different parcellation (Schaefer et al., 2018) for the sustained attention models (see Supplementary Figure A4 and Table A5). However, the interaction effects with mind wandering did not reach significance (see Supplementary Figure A4 and Table A6. In this regard, results our results should be replicated with a larger sample size than ours (n=29).

Conclusion and future directions

The present results could be relevant for clinical disorders with abnormal sustained attention and elevated mind wandering frequency, i.e., in ADHD (Barkley, 1997 and Bozhilova et al., 2018 respectively). In fact, it has been argued that sustained attention could serve as a global summary metric for a subject's general attention abilities, given that sustained attention involves a multi-faceted set of attentional functions, such as alertness, goal directed attention, enhancement of selected information (Coste et al., 2011) and inhibition of task-irrelevant information (Chun et al., 2011). There is evidence that impairments of attention in ADHD and healthy subjects are related to a variety of distributed connections between nodes or networks across the whole-brain connectome (Castellanos & Proal, 2012; Rosenberg et al., 2016). Instead of mapping node-to-node connectivity, our approach was to evaluate how global attributes of brain communication, thought to reflect different modes of information processing, underlies cognitive fluctuations. In sum, this work mapped two converging communication modes throughout the large-scale functional architecture to behavioral fluctuations in sustained attention and task-unrelated thought. This study has the potential to advance current neurocognitive models of attention and establishes a new methodological and theoretical approach to linking brain and behavior.

Acknowledgments

AZ is supported by the Swiss National Science Foundation (P2ZHP1_181435). CMW is supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research

Foundation) under Germanys Excellence Strategy – EXC 2064/1 – 390727645. This research was supported by a Merit Review Award from the Department of Veterans Affairs Clinical Sciences Research and Development awarded to M.E. (I01CX001653).

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, *8*, 14.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of t he brain's default network. *Neuron*, *65*(4), 550–562.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of adhd. *Psychological bulletin*, *121*(1), 65.

Bassett, D. S., Yang, M., Wymbs, N. F., & Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature neuroscience*, *18*(5), 744–751.

Bastian, M., & Sackur, J. (2013). Mind wandering at the fingertips: Automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, *4*, 573.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 1165–1188.

Bertolero, M. A., Yeo, B. T., & DEsposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, *112*(49), E6798–E6807.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Bozhilova, N. S., Michelini, G., Kuntsi, J., & Asherson, P. (2018). Mind wandering perspective on attention-deficit/hyperactivity disorder. *Neuroscience & Biobehavioral Reviews*, *92*, 464–476.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J. I., Grimm, O., Heinz, A., Tost, H., Et al. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, (37), 11678–11683.

Castellanos, F. X., & Proal, E. (2012). Large-scale brain systems in adhd: Beyond the prefrontal–striatal model. *Trends in cognitive sciences*, *16*(1), 17–26.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fmri reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*(21), 8719–8724.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, *17*(11), 718.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, *62*, 73–101.

Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, *36*(48), 12083–12094.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature neuroscience*, *16*(9), 1348–1355.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201–215.

Coste, C. P., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2011). Ongoing brain activity fluctuations directly account for intertrial and indirectly for intersubject variability in stroop task performance. *Cerebral cortex*, *21*(11), 2612–2619.

Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., McGuire, P., & Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, *110*(28), 11583–11588.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529–14534.

Ekman, M., Derrfuss, J., Tittgemeyer, M., & Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences*, *109*(41), 16714–16719.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Et al. (2019). Fmriprep:

A robust preprocessing pipeline for functional mri. *Nature methods*, *16*(1), 111–116.

Esterman, M., Grosso, M., Liu, G., Mitko, A., Morris, R., & DeGutis, J. (2016). Anticipation of monetary reward can attenuate the vigilance decrement. *PloS one*, *11*(7), e0159741.

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723.

Esterman, M., Rosenberg, M. D., & Noonan, S. K. (2014). Intrinsic fluctuations in sustained attention and distractor processing. *Journal of Neuroscience*, *34*(5), 1724–1730.

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*.

Esterman, M., Thai, M., Okabe, H., DeGutis, J., Saad, E., Laganiere, S. E., & Halko, M. A. (2017). Network-targeted cerebellar transcranial magnetic stimulation improves attentional control. *NeuroImage*, *156*, 190–198.

Finc, K., Bonna, K., He, X., Lydon-Staley, D. M., Kühn, S., Duch, W., & Bassett, D. S. (2020). Dynamic reconfiguration of functional brain networks during working memory training. *Nature Communications*, *11*(1), 1–15.

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*, *26*(9), 1497–1510.

Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J., & Esterman, M. (2018). Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension. *Neuroimage*, *171*, 148–164.

Fortenbaugh, F. C., Silver, M., & Robertson, L. (2013). Redefining the metric of visual space: Visual field boundaries influence attentional resolution and crowding performance. *Journal of Vision*, *13*(9), 577–577.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75–174.

Fox, K. C., Spreng, R. N., Ellamil, M., Andrews-Hanna, J. R., & Christoff, K. (2015). The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. *Neuroimage*, *111*, 611–621.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.

Godwin, C. A., Hunter, M. A., Bezdek, M. A., Lieberman, G., Elkin-Frankston, S., Romero, V. L., Witkiewitz, K., Clark, V. P., & Schumacher, E. H. (2017). Functional connectivity within and between intrinsic brain networks correlates with trait mind wandering. *Neuropsychologia*, *103*, 140–153.

Godwin, D., Barry, R. L., & Marois, R. (2015). Breakdown of the brains functional network modularity with awareness. *Proceedings of the National Academy of Sciences*, *112*(12), 3799–3804.

Golchert, J., Smallwood, J., Jefferies, E., Seli, P., Huntenburg, J. M., Liem, F., Lauckner, M. E., Oligschläger, S., Bernhardt, B. C., Villringer, A., Et al. (2017). Individual variation in intentionality in the mind-wandering state is reflected in the integration of the default-mode, fronto-parietal, and limbic networks. *Neuroimage*, *146*, 226–235.

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences*, *112*(44), 13681–13686.

Guimera, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature*, *433*(7028), 895–900.

Jamalabadi, H., Zuberer, A., Kumar, V. J., Li, M., Alizadeh, S., Moradi, A. A., Gaser, C., Esterman, M., & Walter, M. (2020). The missing role of gray matter in studying brain controllability. *bioRxiv*.

Kucyi, A., & Davis, K. D. (2014). Dynamic functional connectivity of the default mode network tracks daydreaming. *Neuroimage*, *100*, 471–480.

Kucyi, A., Esterman, M., Riley, C. S., & Valera, E. M. (2016). Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences*, *113*(48), 13899–13904.

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral cortex*, *27*(3), 1831–1840.

Kucyi, A., Tambini, A., Sadaghiani, S., Keilholz, S., & Cohen, J. R. (2018). Spontaneous cognitive processes and the behavioral validation of time-varying brain connectivity. *Network Neuroscience*, *2*(4), 397–417.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control

task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *Journal of Neuroscience*, *34*(49), 16286–16295.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, *103*(23), 8577–8582.

OCallaghan, C., Shine, J. M., Hodges, J. R., Andrews-Hanna, J. R., & Irish, M. (2019). Hippocampal atrophy and intrinsic brain network dysfunction relate to alterations in mind wandering in neurodegeneration. *Proceedings of the National Academy of Sciences*, *116*(8), 3316–3321.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annu. Rev. Psychol.*, *58*, 1–23.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., Et al. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fmri. *Neuroimage*, *84*, 320–341.

Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N., & Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron*, *79*(4), 798–813.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165.

Rothlein, D., DeGutis, J., & Esterman, M. (2018). Attentional fluctuations influence the neural fidelity and connectivity of stimulus representations. *Journal of cognitive neuroscience*, *30*(9), 1209–1228.

Rubinov, M., Kötter, R., Hagmann, P., & Sporns, O. (2009). Brain connectivity toolbox: A collection of complex network measurements and brain connectivity datasets. *NeuroImage*, (47), S169.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, *52*(3), 1059–1069.

Sadaghiani, S., & D'Esposito, M. (2015). Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*, *25*(9), 2763–2773.

Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., & DEsposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences*, *112*(27), 8463–8468.

Sato, J. R., Takahashi, D. Y., Hoexter, M. Q., Massirer, K. B., & Fujita, A. (2013). Measuring network's entropy in adhd: A new approach to investigate neuropsychiatric disorders. *Neuroimage*, *77*, 44–51.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, *28*(9), 3095–3114.

Seli, P., Cheyne, J. A., Xu, M., Purdon, C., & Smilek, D. (2015). Motivation, intentionality, and mind wandering: Implications for assessments of task-unrelated thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1417.

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The dynamics of functional brain networks: Integrated network states during cognitive task performance. *Neuron*, *92*(2), 544–554.

Shine, J. M., & Poldrack, R. A. (2018). Principles of dynamic network reconfiguration across diverse brain states. *NeuroImage*, *180*, 396–405.

Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., & Snyder, A. Z. (2017). Data quality influences observed links between functional connectivity and behavior. *Cerebral cortex*, *27*(9), 4492–4502.

Smallwood, J., & Schooler, J. W. (2006). The Restless Mind. *Psychological bulletin*, *132*(6), 946.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., Et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, *106*(31), 13040–13045.

Sormaz, M., Murphy, C., Wang, H.-t., Hymers, M., Karapanagiotidis, T., Poerio, G., Margulies, D. S., Jefferies, E., & Smallwood, J. (2018). Default mode network can support the level of detail in experience during active task states. *Proceedings of the National Academy of Sciences*, *115*(37), 9318–9323.

Stanley, M. L., Simpson, S. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2015). Changes in brain network efficiency and working memory performance in aging. *PLoS One*, *10*(4), e0123950.

Stawarczyk, D., Majerus, S., Maquet, P., & D'Argembeau, A. (2011). Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-

independence are related to default network activity. *PloS one*, *6*(2), e16997.

Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PloS one*, *7*(1).

Stuss, D. T. (2006). Frontal lobes and attention: Processes and networks, fractionation and integration. *Journal of the International Neuropsychological Society*, *12*(2), 261–271.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400–410.

Thompson, W. H., Kastrati, G., Finc, K., Wright, J., Shine, J. M., & Poldrack, R. A. (2020). Time-varying nodal measures with temporal community structure: A cautionary note to avoid misinterpretation. *Human Brain Mapping*, *41*(9), 2347–2356.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, *15*(1), 273–289.

Weissman, D. H., Roberts, K., Visscher, K., & Woldorff, M. (2006). The neural bases of momentary lapses in attention. *Nature neuroscience*, *9*(7), 971–978.

Yeo, B. T., Krienen, F. M., Eickhoff, S. B., Yaakub, S. N., Fox, P. T., Buckner, R. L., Asplund, C. L., & Chee, M. W. (2015). Functional specialization and flexibility in human association cortex. *Cerebral cortex*, *25*(10), 3654–3672.

Yue, Q., Martin, R. C., Fischer-Baum, S., Ramos-Nuñez, A. I., Ye, F., & Deem, M. W. (2017). Brain modularity mediates the relation between task complexity and performance. *Journal of cognitive neuroscience*, *29*(9), 1532–1546.

Zimmer, U., Roberts, K. C., Harshbarger, T. B., & Woldorff, M. G. (2010). Multisensory conflict modulates the spread of visual attention across a multisensory object. *Neuroimage*, *52*(2), 606–616.

Appendix
Supplementary materials

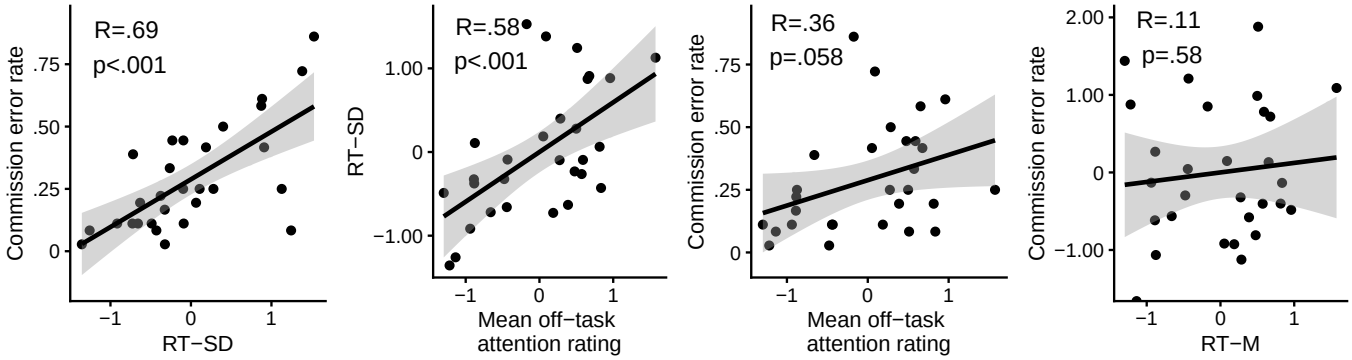

Figure A1
Behavioral results. Pearson correlation of RT-SD with attention lapses (commission errors), mean off-task rating and RT-SD, and mean off-task rating with attention lapses (commission errors), and mean reaction time (RT-M) with attention lapses (commission errors). RT-SD, standard deviation of reaction time, RT-M, mean of reaction time

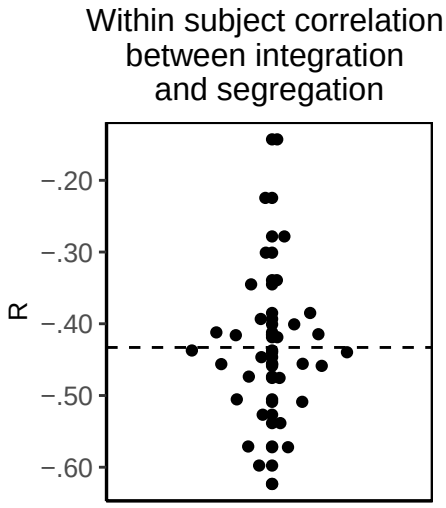

Figure A2
Within-subject Pearson correlation between integration and segregation. The mean within-subject correlation was negative ($r = -0.43$) and significantly greater smaller zero (T_{28}, $p < 0.001$)

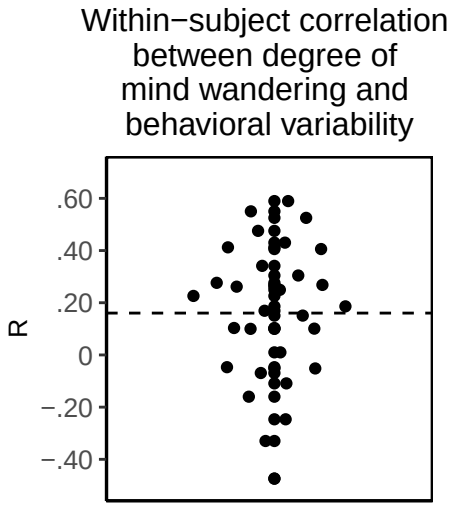

Figure A3

Within-subject Pearson correlation between degree of mind wandering and behavioral variability The mean within-subject correlation was positive ($r = 0.16$) and significantly greater than zero (T_{28}, $p = .003$).

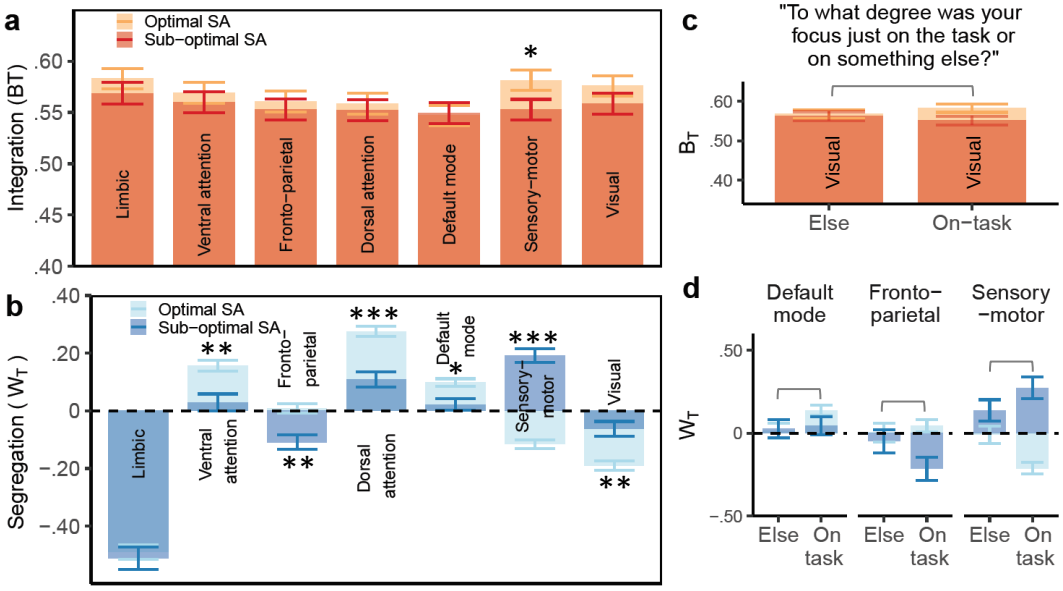

Figure A4

Sustained attention network (SAN) model based on two modes of information processing across the whole-brain connectome, the diversity of cross-community connectivity (B_T) and degree of within-community connectivity (W_T). Network assignments are derived from the Schaefer parcellation with 200 regions (Schaefer et al., 2018). **(a)** and **(b)** display the interaction effect of RT variability (dichotomized into optimal and sub-optimal sustained attention, corresponding to low and high variability, respectively) and network assignment for a linear mixed effects model predicting B_T and W_T. The dichotomization of RT variability (performed using a median split) was only for visualization purposes, whereas the original models used a continuous variable. **(c)** and **(d)** display the interaction effect of RT variability, degree of mind wandering, and network assignment for a linear. Significance levels: $*p < .05$, $**p < .01$, $***p < .001$.

Table A1

Mixed Effects Regression. Sustained attention model.

	B_T			W_T		
Predictors	Estimates	CI	*p*	Estimates	CI	*p*
(Intercept)	.03	-.26 - .33	.815	.73	.71 - .75	<.001
Au	-.03	-.07 - .00	.086	-1.49	-1.52 - -1.46	<.001
Ce	.23	.19 - .27	<.001	-2.82	-2.85 - -2.79	<.001
CO	-.29	-.33 - -.26	<.001	-1.02	-1.05 - -.99	<.001
DA	-.47	-.51 - -.43	<.001	-1	-1.03 - -.97	<.001
DM	-.19	-.23 - -.15	<.001	-1.04	-1.07 - -1.01	<.001
FP	-.01	-.04 - .03	.771	-1.72	-1.75 - -1.69	<.001
Sa	-.04	-.08 - -.00	.039	-1.41	-1.44 - -1.38	<.001
SM	-.03	-.06 - -.00	.042	-1.41	-1.43 - -1.39	<.001
Su	.39	.35 - .43	<.001	-2.52	-2.55 - -2.49	<.001
VA	.2	.16 - .23	<.001	-1.87	-1.90 - -1.84	<.001
Vi	-.2	-.24 - -.17	<.001	-1.18	-1.21 - -1.15	<.001
RTsd	-.01	-.02 - .00	.143	0	-.01 - .00	.332
run	-.01	-.02 - -.01	<.001			
Au * RTsd	-.09	-.13 - -.06	<.001	.26	.23 - .29	<.001
Ce * RTsd	-.01	-.05 - .03	.551	.02	-.01 - .05	.187
CO * RTsd	.13	.09 - .16	<.001	-.13	-.16 - -.10	<.001
DA * RTsd	.17	.13 - .20	<.001	-.17	-.20 - -.14	<.001
DM * RTsd	0	-.03 - .04	.859	-.07	-.10 - -.04	<.001
FP * RTsd	0	-.04 - .03	.912	-.01	-.04 - .02	.35
Sa * RTsd	.06	.02 - .10	.001	-.07	-.10 - -.04	<.001
SM * RTsd	-.09	-.12 - -.07	<.001	.18	.16 - .20	<.001
Su * RTsd	-.02	-.06 - .02	.341	.07	.04 - .10	<.001
VA * RTsd	-.02	-.06 - .01	.203	.06	.03 - .09	<.001
Vi * RTsd	.07	.03 - .10	<.001	-.05	-.08 - -.02	.002
Random Effects						
σ^2	.35			.22		
τ_{00}	.64			.00		
N	29 Subjects			29 Subjects		
Marginal R^2	.028			.773		
Conditional R^2	.655			.776		

Note: B_T is the diversity of connectivity across time-varying communities (quantified by the participation coefficient), W_T is the degree of connectivity within time-varying communities (quantified by within-module degree Z score). Network assignment was derived from the Power parcellation (Power et al., 2011). Au, Auditory; Ce, Cerebellar; CO, Cingulo-opercular; DM, Default mode; DA, Dorsal attention; FP, Fronto-parietal; Sa, Salience; SM, sensory/somato-motor; VA, Ventral attention; Vi, Visual; RTsd, standard deviation of reaction time; σ^2, random effect variance; τ_{00}, between-subject variance; CI, bootstrapped 95% confidence intervals. All *p*-values were computed via Wald-statistics approximation (treating *t* as Wald *z*).

Table A2

Mixed Effects Regression. Sustained attention model interacting with task-unrelated thought.

	B_T			W_T		
Predictors	Estimates	CI	p	Estimates	CI	p
Intercept	.03	-.26 .33	.822	.73	.71 .75	<.001
Au	-.03	-.07 .01	.135	-1.48	-1.51 -1.45	<.001
Ce	.23	.19 .27	<.001	-2.83	-2.86 -2.79	<.001
CO	-.3	-.34 -.26	<.001	-1.01	-1.05 -.98	<.001
DA	-.46	-.50 -.42	<.001	-1	-1.03 -.96	<.001
DMN	-.19	-.23 -.15	<.001	-1.06	-1.09 -1.03	<.001
FP	-.01	-.05 .04	.792	-1.73	-1.77 -1.70	<.001
Sa	-.05	-.09 -.01	.015	-1.41	-1.44 -1.37	<.001
SM	-.02	-.05 .01	.111	-1.4	-1.43 -1.38	<.001
Su	.39	.35 .43	<.001	-2.5	-2.53 -2.47	<.001
VA	.2	.16 .24	<.001	-1.87	-1.90 -1.83	<.001
Vi	-.19	-.23 -.15	<.001	-1.19	-1.22 -1.16	<.001
RTsd	-.01	-.02 .01	.278	-.01	-.02 .01	.328
TF	-.01	-.02 .00	.133	0	-.01 .01	.47
Run	-.01	-.02 -.00	.001			
Au * RTsd	-.08	-.12 -.04	<.001	.25	.22 .28	<.001
Ce * RTsd	-.02	-.06 .02	.37	-.02	-.05 .02	.308
CO * RTsd	.13	.09 .17	<.001	-.13	-.16 -.10	<.001
DA * RTsd	.14	.10 .18	<.001	-.11	-.14 -.08	<.001
DMN * RTsd	-.01	-.06 .03	.482	-.06	-.09 -.03	<.001
FP * RTsd	-.01	-.05 .03	.64	.01	-.03 .04	.738
Sa * RTsd	.06	.02 .10	.005	-.07	-.10 -.04	<.001
SM * RTsd	-.08	-.11 -.05	<.001	.16	.13 .18	<.001
Su * RTsd	-.02	-.06 .02	.446	.07	.04 .11	<.001
VA * RTsd	-.01	-.06 .03	.48	.05	.02 .08	.003
Vi * RTsd	.06	.02 .10	.003	-.05	-.08 -.01	.004
Au * TF	-.03	-.07 .01	.15	.04	.01 .07	.016
Ce * TF	.02	-.02 .06	.384	.09	.06 .12	<.001
CO * TF	0	-.04 .04	.911	.01	-.02 .04	.514
DA * TF	.08	.03 .12	<.001	-.14	-.17 -.11	<.001
DMN * TF	.05	.00 .09	.031	-.03	-.06 .00	.05
FP * TF	.02	-.02 .06	.347	-.06	-.09 -.02	.001
Sa * TF	0	-.04 .04	.898	-.01	-.04 .03	.705
SM * TF	-.04	-.07 -.01	.004	.06	.03 .08	<.001
Su * TF	-.01	-.05 .03	.763	-.01	-.04 .02	.649
VA * TF	-.02	-.06 .02	.315	.03	-.00 .06	.054
Vi * TF	.02	-.02 .06	.328	-.01	-.04 .03	.75
RTsd * TF	0	-.01 .01	.911	0	-.01 .01	.673
Au * RTsd * TF	0	-.04 .04	.967	-.03	-.06 -.00	.046
Ce * RTsd * TF	0	-.04 .04	.947	0	-.03 .03	.974
CO * RTsd * TF	.01	-.03 .05	.646	-.02	-.05 .01	.134
DA * RTsd * TF	-.03	-.06 .01	.166	.02	-.01 .05	.276
DMN * RTsd * TF	0	-.04 .04	.978	.04	.01 .07	.01
FP * RTsd * TF	0	-.04 .04	.901	.04	.01 .07	.015
Sa * RTsd * TF	.03	-.01 .06	.165	-.01	-.05 .02	.331
SM * RTsd * TF	-.01	-.03 .02	.71	-.03	-.05 -.01	.005
Su * RTsd * TF	0	-.03 .04	.831	-.05	-.08 -.02	<.001
VA * RTsd * TF	-.01	-.05 .03	.532	-.02	-.05 .01	.109
Vi * RTsd * TF	-.04	-.08 -.00	.034	.02	-.01 .05	.216
Random Effects						
σ^2	.35			.22		
$\tau_0 0$	.64			.00		
N	29 Subjects			29 Subjects		
Marginal R^2	.028			.775		
Conditional R^2	.656			0.777		

Note: B_T is the diversity of connectivity across time-varying communities (quantified by the participation coefficient), W_T is the degree of connectivity within time-varying communities (quantified by within-module degree Z score). Network assignment was derived from the Power parcellation (Power et al., 2011). Au, Auditory; Ce, Cerebellar; CO, Cingulo-opercular; DMN, Default mode; DA, Dorsal attention; FP, Fronto-parietal; SA, Salience; SM, sensory/somato-motor; VA, Ventral attention; Vi, Visual; RTsd, standard deviation of reaction time; TF, task focus. σ^2, random effect variance; τ_{00}, between-subject variance; CI, bootstrapped 95% confidence intervals. All p-values were computed via Wald-statistics approximation (treating t as Wald z).

Table A3

Rich hubs of optimal sustained attention

Node index	Node label	Putative system	Estimate W_T	Estimate B_T	MNI x	MNI y	MNI z
54	Supp_Motor_Area_R	Cingulo-opercular Task Control	-.0192	.0130	7	8	51
251	Precuneus_R	Dorsal attention	-.0171	.0104	10	-62	61
99	Frontal_Sup_L	Default mode	-.0145	.0059	-16	29	53
104	Frontal_Sup_L	Default mode	-.0143	.0066	-20	45	39
170	Calcarine_R	Visual	-.0141	.0113	6	-81	6
105	Frontal_Sup_Medial_R	Default mode	-.0141	.0055	6	54	16
47	Supp_Motor_Area_L	Cingulo-opercular Task Control	-.0133	.0110	-3	2	53
146	Calcarine_L	Visual	-.0116	.0095	-8	-81	7
27	Postcentral_L	Sensory/somatomotor Hand	-.0109	.0095	-38	-27	69
137	Frontal_Inf_Orb_L	Default mode	-.0108	.0047	-46	31	-13
263	Parietal_Sup_L	Dorsal attention	-.0107	.0074	-17	-59	64
139	Frontal_Inf_Orb_R	Default mode	-.0104	.0042	49	35	-12
145	Calcarine_R	Visual	-.0104	.0099	8	-72	11
103	Frontal_Sup_L	Default mode	-.0096	.0065	-10	55	39
213	Supp_Motor_Area_L	Salience	-.0096	.0066	-1	15	44
26	Postcentral_R	Sensory/somatomotor Hand	-.0095	.0074	50	-20	42
83	Temporal_Mid_L	Default mode	-.0086	.0042	-68	-23	-16
37	undefined	Sensory/somatomotor Hand	-.0084	.0084	-38	-15	69
255	Postcentral_R	Sensory/somatomotor Hand	-.0081	.0099	47	-30	49
209	Insula_R	Salience	-.0079	.0083	36	22	3
50	Frontal_Sup_L	Cingulo-opercular Task Control	-.0077	.0047	-16	-5	71

Note: Table shows hub nodes of optimal attention exhibiting concurrently high segregation and low integration. Node index, index of node of the Power parcellation (Power et al., 2011), Node label, labels derived from the Automated Anatomical Labeling (Tzourio-Mazoyer et al., 2002) fitted onto the Power atlas. Estimate W_T, Estimate of interaction term between reaction time variability and node index of the Power atlas derived from a mixed effects model predicting segregation of information processing, quantified by the within-module degree z-score, Estimate B_T, Estimate of interaction term between reaction time variability and node index of the Power parcellation derived from a mixed effects model predicting integration of information processing, quantified by the participation coefficient, MNI *xyz*, coordinates of node position in MNI space.

Table A4

Diverse hub nodes of optimal sustained attention

Node index	Node label	Putative system	Estimate B_T	Estimate W_T	MNI x	MNI y	MNI z
63	Temporal_Sup_R	Auditory	-.0085	.0192	58	-16	7
46	Postcentral_R	Sensory/somatomotor Mouth	-.0084	.0158	66	-8	25
42	Postcentral_L	Sensory/somatomotor Mouth	-.0075	.0088	-49	-11	35
41	Precentral_R	Sensory/somatomotor Hand	-.0072	.0130	38	-17	45
19	Postcentral_R	Sensory/somatomotor Hand	-.0071	.0105	13	-33	75
66	Temporal_Sup_L	Auditory	-.0068	.0132	-49	-26	5
45	Postcentral_L	Sensory/somatomotor Mouth	-.0068	.0071	-53	-10	24
31	Supp_Motor_Area_R	Sensory/somatomotor Hand	-.0067	.0095	10	-17	74
18	Paracentral_Lobule_L	Sensory/somatomotor Hand	-.0067	.0097	-7	-33	72
44	Postcentral_R	Sensory/somatomotor Mouth	-.0065	.0101	51	-6	32
237	Temporal_Sup_L	Ventral attention	-.0064	.0060	-55	-40	14
39	Paracentral_Lobule_R	Sensory/somatomotor Hand	-.0062	.0095	2	-28	60
71	Rolandic_Oper_R	Auditory	-.0061	.0143	56	-5	13
90	Precuneus_L	Default mode	-.0056	.0072	-11	-56	16
28	undefined	Sensory/somatomotor Hand	-.0056	.0087	20	-29	60
62	Temporal_Sup_R	Auditory	-.0051	.0133	65	-33	20
70	Rolandic_Oper_L	Auditory	-.0051	.0090	-55	-9	12
61	undefined	Auditory	-.0050	.0077	32	-26	13
5	Rectus_R	Uncertain	-.0050	.0060	8	41	-24
36	Precentral_R	Sensory/somatomotor Hand	-.0046	.0072	42	-20	55
17	Paracentral_Lobule_L	Sensory/somatomotor Hand	-.0046	.0066	-7	-21	65
117	Temporal_Mid_L	Default mode	-.0045	.0073	-56	-13	-10
43	Insula_R	Sensory/somatomotor Mouth	-.0045	.0101	36	-9	14
192	Parietal_Inf_R	Fronto-parietal Task Control	-.0042	.0089	44	-53	47
141	Lingual_R	Uncertain	-.0040	.0056	17	-91	-14
64	Rolandic_Oper_L	Auditory	-.0039	.0148	-38	-33	17

Note: Node index, index of node of the power atlas (Power et al., 2011), Table shows hub nodes of optimal attention exhibiting concurrently high integration and low segregation. Node label, labels derived from the Automated Anatomical Labeling (Tzourio-Mazoyer et al., 2002) fitted onto the Power atlas. Estimate B_T, Estimate of interaction term between reaction time variability and node index of the Power atlas derived from a mixed effects model predicting integration of information processing, quantified by the participation coefficient,Estimate W_T, Estimate of interaction term between reaction time variability and node index of the Power atlas derived from a mixed effects model predicting segregation of information processing, quantified by the within-module degree z-score, MNI *xyz*, coordinates of node position in MNI space.

Table A5

Mixed Effects Regression. Sustained attention model (Schaefer parcellation).

	B_T			W_T		
Predictors	Estimates	CI	p	Estimates	CI	p
Intercept	-.08	-.33 .17	.54	.13	.09 .17	<.001
DA	-.13	-.19 -.06	<.001	.78	.72 .84	<.001
DM	-.33	-.39 -.27	<.001	.25	.19 .31	<.001
FP	-.09	-.15 -.02	.006	-.14	-.20 -.08	<.001
Li	.41	.35 .47	<.001	-1.86	-1.91 -1.80	<.001
SM	.22	.16 .29	<.001	.03	-.03 .08	.34
VA	.11	.05 .18	<.001	.4	.34 .45	<.001
Vi	.2	.14 .26	<.001	-.52	-.57 -.46	<.001
RTsd	-.04	-.09 .01	.08	0	-.04 .04	.988
Run	.01	.00 .03	.037			
DA * RTsd	.02	-.04 .08	.524	-.12	-.18 -.07	<.001
DM * RTsd	.05	-.01 .12	.081	-.06	-.11 -.00	.045
FP * RTsd	.01	-.05 .07	.659	-.09	-.14 -.03	.002
Li * RTsd	-.01	-.08 .05	.656	-.02	-.07 .04	.58
SM * RTsd	-.07	-.14 -.01	.016	.23	.17 .28	<.001
VA * RTsd	.01	-.05 .07	.802	-.1	-.15 -.04	.001
Vi * RTsd	-.03	-.09 .03	.364	.09	.04 .15	.001
Random Effects						
σ^2	.52			.43		
τ_{00}	.45			.00		
N	29 Subjects			29 Subjects		
Marginal R^2	.052			.568		
Conditional R^2	.496			.569		

Note: B_T is the diversity of connectivity across time-varying communities (quantified by the participation coefficient), W_T is the degree of connectivity within time-varying communities (quantified by within-module degree Z score). Networks were derived from the Schaefer parcellation (Schaefer et al., 2018) with 200 regions. DA, Dorsal attention; DM, Default mode; FP, Fronto-parietal; Li, Limbic; SM, sensory/somato-motor; VA, Ventral attention; Vi, Visual; RTsd, standard deviation of reaction time; σ^2, random effect variance; τ_{00}, between-subject variance; CI, bootstrapped 95% confidence intervals. All p-values were computed via Wald-statistics approximation (treating t as Wald z).

Table A6

Mixed Effects Regression. Sustained attention model interacting with task-unrelated thought (Schaefer parcellation).

	B_T			W_T		
Predictors	Estimates	CI	p	Estimates	CI	p
Intercept	-.11	-.36 .15	.416	.13	.09 .18	<.001
DA	-.13	-.20 -.07	<.001	.8	.74 .86	<.001
DM	-.34	-.40 -.27	<.001	.25	.19 .31	<.001
FP	-.09	-.15 -.02	.011	-.16	-.22 -.09	<.001
Li	.42	.35 .49	<.001	-1.87	-1.93 -1.80	<.001
SM	.22	.16 .29	<.001	.05	-.02 .11	.146
VA	.12	.05 .18	.001	.38	.32 .44	<.001
Vi	.21	.14 .28	<.001	-.52	-.58 -.46	<.001
RTsd	-.05	-.10 .00	.055	0	-.04 .04	.986
TF	-.04	-.09 .01	.095	0	-.04 .04	.99
Run	.02	.01 .03	.005			
DA * RTsd	0	-.07 .07	.972	-.09	-.15 -.03	.006
DM * RTsd	.04	-.02 .11	.207	-.03	-.09 .03	.34
FP * RTsd	.02	-.04 .09	.477	-.11	-.17 -.04	.001
Li * RTsd	.01	-.06 .07	.823	-.05	-.11 .01	.09
SM * RTsd	-.06	-.13 .00	.067	.21	.15 .27	<.001
VA * RTsd	.02	-.05 .08	.637	-.11	-.17 -.05	<.001
Vi * RTsd	-.03	-.10 .03	.344	.09	.03 .15	.005
DA * TF	.05	-.02 .12	.14	-.08	-.15 -.02	.007
DM * TF	.03	-.04 .09	.434	-.07	-.13 -.01	.022
FP * TF	-.03	-.09 .04	.424	.04	-.02 .10	.224
Li * TF	-.05	-.12 .02	.139	.09	.03 .15	.004
SM * TF	-.03	-.10 .04	.363	.06	-.00 .12	.062
VA * TF	-.02	-.09 .05	.542	.04	-.02 .10	.234
Vi * TF	.01	-.05 .08	.664	.01	-.05 .07	.671
RTsd * TF	.05	.00 .09	.041	0	-.04 .04	.992
DA * RTsd * TF	.01	-.05 .07	.754	-.04	-.10 .01	.146
DM * RTsd * TF	.01	-.06 .07	.858	.02	-.04 .07	.597
FP * RTsd * TF	0	-.06 .07	.889	.03	-.03 .09	.292
Li * RTsd * TF	-.01	-.08 .05	.648	.02	-.04 .07	.596
SM * RTsd * TF	0	-.06 .07	.902	-.05	-.11 .01	.087
VA * RTsd * TF	0	-.06 .06	.977	.03	-.03 .09	.294
Vi * RTsd * TF	-.02	-.08 .04	.491	.01	-.04 .07	.671
Random Effects						
σ^2	.51			.43		
τ_{00}	.47			.00		
N	29 Subjects			29 Subjects		
Marginal R^2	.058			.572		
Conditional R^2	.51			.572		

Note: B_T is the diversity of connectivity across time-varying communities (quantified by the participation coefficient), W_T is the degree of connectivity within time-varying communities (quantified by within-module degree Z score). Networks were derived from the Schaefer parcellation (Schaefer et al., 2018) with 200 regions. DA, Dorsal attention; DM, Default mode; FP, Fronto-parietal; Li, Limbic; SM, sensory/somato-motor; VA, Ventral attention; Vi, Visual; RTsd, standard deviation of reaction time; TF, task focus. σ^2, random effect variance; τ_{00}, between-subject variance; CI, bootstrapped 95% confidence intervals. All p-values were computed via Wald-statistics approximation (treating t as Wald z).