



OPEN

Variable rather than extreme slow reaction times distinguish brain states during sustained attention

Ayumu Yamashita^{1,2,3}✉, David Rothlein^{1,2}, Aaron Kucyi⁴, Eve M. Valera^{5,6}, Laura Germine^{5,7}, Jeremy Wilmer⁸, Joseph DeGutis^{2,5,9} & Michael Esterman^{1,2,9}

A common behavioral marker of optimal attention focus is faster responses or reduced response variability. Our previous study found two dominant brain states during sustained attention, and these states differed in their behavioral accuracy and reaction time (RT) variability. However, RT distributions are often positively skewed with a long tail (i.e., reflecting occasional slow responses). Therefore, a larger RT variance could also be explained by this long tail rather than the variance around an assumed normal distribution (i.e., reflecting pervasive response instability based on both faster and slower responses). Resolving this ambiguity is important for better understanding mechanisms of sustained attention. Here, using a large dataset of over 20,000 participants who performed a sustained attention task, we first demonstrated the utility of the exGaussian distribution that can decompose RTs into a strategy factor, a variance factor, and a long tail factor. We then investigated which factor(s) differed between the two brain states using fMRI. Across two independent datasets, results indicate unambiguously that the variance factor differs between the two dominant brain states. These findings indicate that 'suboptimal' is different from 'slow' at the behavior and neural level, and have implications for theoretically and methodologically guiding future sustained attention research.

Although sustaining attention over time is important for our everyday life, evidence suggests our attention fluctuates from moment to moment despite our efforts to maintain optimal focus. Previous studies have found significant individual differences in sustained attention ability in the healthy population, and have related these differences to real world outcomes such as driving accidents, school performance, and memory encoding^{1–5}. Furthermore, impairments in sustained attention are common in a diverse range of clinical populations such as attention-deficit/hyperactivity disorder (ADHD)^{6–8}, major depressive disorder⁹, schizophrenia¹⁰, bipolar disorder¹¹, post-traumatic stress disorder^{12–15}, early-life trauma¹⁶, and traumatic brain injury¹⁷. Therefore, revealing the neural and cognitive mechanisms required to sustain optimal attention is an important issue in modern society.

To better characterize sustained attention, attentional states have been behaviorally estimated using response accuracy^{17–19} and reaction time (RT) to stimuli^{6,20–23}. In particular, RT is frequently used in a variety of cognitive tasks regardless of the difficulty of the cognitive tasks. For example, previous studies found that fast RTs frequently occur before commission errors, thought to reflect lapses due to impulsiveness or mindlessness. This interpretation is also consistent with the result which shows that fast RTs are associated with task-unrelated thoughts^{17,20,23}. On the other hand, slow RTs frequently occur before omission errors, thought to reflect more severe attentional lapses due to task disengagement²⁰. Therefore, both fast RTs, which are interpreted as impulsivity and mindless automatic responding, and slow RTs, which are interpreted as disengaged/inefficient processing of stimulus information, are related to inattention. As an alternative to the speed of RT, intra-individual variability of RT has been recognized as an important indicator of sustained attention^{21,24,25}. Attention fluctuates between stable/accurate (in-the-zone) and variable/error-prone (out-of-the-zone) states^{21,26}. Furthermore, one of the most consistent manifestations of ADHD is the high prevalence of intra-individual RT variability^{6,8}.

¹Department of Psychiatry, Boston University School of Medicine, Boston, MA 02118, USA. ²Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA 02130, USA. ³Overseas Research Fellow, Japan Society for the Promotion of Science, Tokyo 102-0083, Japan. ⁴Department of Psychology, Northeastern University, Boston, MA 02115, USA. ⁵Department of Psychiatry, Harvard Medical School, Massachusetts 02215, USA. ⁶Department of Psychiatry, Massachusetts General Hospital, Massachusetts 02129, USA. ⁷Institute for Technology in Psychiatry, McLean Hospital, Boston, MA 02478, USA. ⁸Wellesley College, Wellesley, MA 02481, USA. ⁹National Center for PTSD, VA Boston Healthcare System, Boston, MA 02130, USA. ✉email: ayamashi@bu.edu

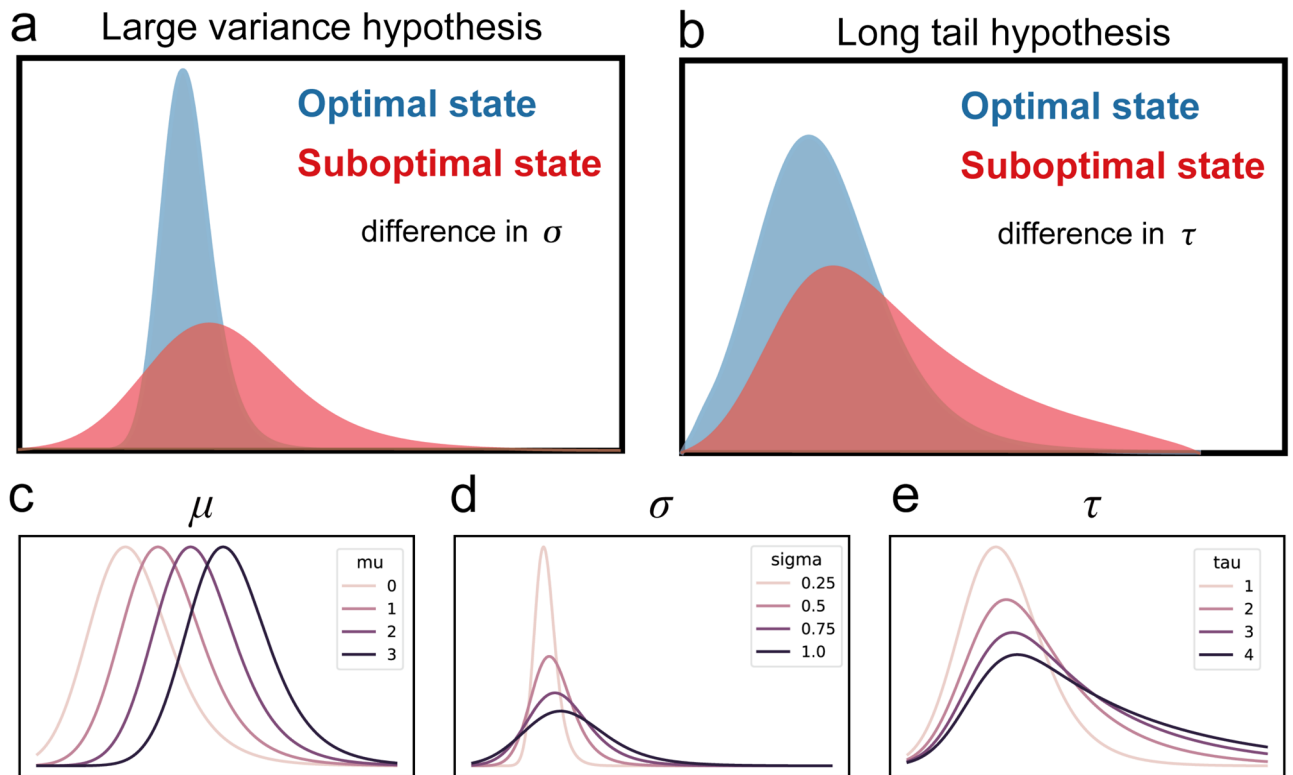


Figure 1. Two alternative hypotheses and exGaussian distributions. (a) Illustration of large variance hypothesis. (b) Illustration of long tail hypothesis. (c) exGaussian distributions with different μ ($\mu = \{0, 1, 2, 3\}$, $\sigma = 1$, $\tau = 1$). (d) exGaussian distributions with different σ ($\mu = 0$, $\sigma = \{0.25, 0.50, 0.75, 1.0\}$, $\tau = 1$). (e) exGaussian distributions with different τ ($\mu = 0$, $\sigma = 1$, $\tau = \{1, 2, 3, 4\}$).

From the viewpoint of the RT distribution across task trials, the mean of the RTs represents the response speed while the variance of the RTs represents the intra-individual variability. Sustained attention tasks vary considerably. They include simple reaction time tasks such as the psychomotor vigilance task (PVT), which do not have a response deadline, and continuous performance tasks (CPT) such as sustained attention to response task (SART) and gradual onset CPT (gradCPT), which have an implicit deadline for a response as the trials follow in rapid succession. Given this implicit response deadline in CPTs, which does not allow extremely slow responses, mean and variance had been the main focus of research using these tools. While RT mean and variance are intuitive and useful, these parameters are usually inappropriate because, despite response deadlines, RT distributions rarely conform to Gaussian distribution. RT distributions are typically positively skewed with a long tail indicating more extreme slow responses than would be found in a Gaussian distribution. That is, a large RT variance could be fully accounted for by the long tail factor of the RT distribution. Hence, we believe that distinguishing the distributional drivers of RT variability is important to better tease apart the information contained in RT distributions and their relationship to sustained attention. The exGaussian distribution is commonly used to decompose RTs into μ and σ , that, respectively, describe the mean and standard deviation (variability) of the Gaussian distribution, and an independent exponentially distributed variable, τ , that accounts for the positive skew (Fig. 1c–e)^{27,28}. Theoretically, $mean\ RT = \mu + \tau$, and $RT\ variance = \sigma^2 + \tau^2$. Among the three ex-Gaussian parameters, a previous study of sustained attention in ADHD in adolescents showed τ was specifically positively associated with omission errors, suggesting more frequent disengagement-based attention lapses; μ was significantly negatively associated with commission errors suggesting an impulsive, or mindless response style²⁹. Further, this study and others found that ADHD was not associated with generalized RT slowing, which is characterized by μ , but rather was associated with both the long tail factor of slower RT, which is characterized by τ , and with the variance factor, which is characterized by σ ^{29,30}. However, there are still only a few studies applying exGaussian distribution to RTs in sustained attention tasks such as continuous performance task^{22,29–31}. Therefore, in the first part of this study, we examined the relationship between exGaussian parameters and sustained attention performance using a big dataset including over 20,000 participants.

In the second part of this study, we used these tools to investigate the neural mechanism of sustained attention. In our previous study, we observed two dominant brain states during sustained attention task using functional magnetic resonance imaging (fMRI)³². One state (State1) was characterized by activation of frontoparietal network subsystem A (FPN_A), default mode network (DMN) and limbic network, and the other state (State2) was characterized as activation of frontoparietal network subsystem B (FPN_B), dorsal attention network (DAN), salience network (SN), somato-motor (SMN) and visual network. RT variance was lower and accuracy was higher during State1 than during State2, suggesting State1 was optimal and State2 was suboptimal. However,

since this study assumed a Gaussian RT distribution, it is not known whether the difference in RT variance between brain states was due to the variance factor or the long tail factor. Determining which factor(s) differ between brain states is important for better understanding the mechanisms of sustained attention. For example, differences in the variance factor, on the one hand, suggests attentional/brain states reflects variability across the entire empirical RT distribution, including both impulsive fast RT and slow RTs (Large variance hypothesis; Fig. 1a). Previous studies suggest this factor may represent exploration in the environment to gather information or adjust cognitive control settings³³. On the other hand, differences in the long tail factor suggests attentional/brain states reflects only more infrequent and extremely slow RTs suggesting occasional attention lapses (Long tail hypothesis; Fig. 1b)^{22,29,34}. Here, by comparing exGaussian parameters between these two brain states, we can help resolve this ambiguity present in many studies, including our recent work. For the purpose of replication, we tested each of these two hypotheses using two different independent fMRI datasets.

Materials and methods

Overview. We conducted our analyses with three datasets. In all datasets, all participants performed a gradCPT. The gradCPT is a well-validated test of sustained attention, previously used to define attentional states by reaction time variability fluctuations over time^{2,21,26}. Dataset1 consists of gradCPT behavioral data from 21,406 participants between 10 and 70 years old who performed one session of 4 min gradCPT with an interstimulus interval (ISI) of 800 ms obtained via the website testmybrain.org². Dataset1 was used to investigate the relationship between exGaussian parameters and sustained attention performance. Dataset2 consists of gradCPT behavioral data with fMRI from 16 participants who performed three runs of 8 min gradCPT with ISI of 800 ms in MRI scanner²¹. Dataset2 was used to investigate the difference in exGaussian parameters between brain states. Dataset3 consists of gradCPT behavioral data with fMRI from 29 participants who performed four runs with thought probes of 9 min gradCPT with ISI of 1300 ms in MRI scanner³⁵. Dataset3 was used to replicate the results in the Dataset2.

Behavioral analysis. *gradual onset continuous performance task.* The gradCPT contained 10 round, gray-scale photographs of mountain scenes and 10 of city scenes. These scenes were randomly presented with 10% mountain and 90% city trials, without allowing the identical scene to repeat on consecutive trials. Scene images gradually transitioned from one to the next, using a linear pixel-by-pixel interpolation, with each transition occurring in 800 ms (fast version; Dataset1 and Dataset2) or 1300 ms (slow version; Dataset3). Participants were instructed to press a button for each city scene, and withhold responses to mountain scenes. Response accuracy was emphasized without reference to speed. However, given that the next stimulus would replace the current stimulus in 800/1300 ms, a response deadline was implicit in the task. In the Dataset1, the order of presented images was the same for all participants, but in the Dataset2 and 3, the order of presented images was randomized individually. This method is the same as the one used in our previous studies^{21,32}.

Calculation of reaction time. RT was calculated relative to the beginning of each image transition, such that an RT of 800/1300 ms (slow and fast versions) indicates a button press at the moment image n was 100% coherent and not mixed with other images. A shorter RT indicates that the current scene was still in the process of transitioning from the previous, and a longer RT indicates that the current scene was in the process of transitioning to the subsequent scene. So, for example, an RT of 720/1170 ms would be at the moment of 90% image n and 10% image $n - 1$, and so forth. On rare trials with highly deviant RTs (before 70% coherence of image n and after 40% coherence of image $n + 1$) or multiple button presses, an iterative algorithm maximized correct responses as follows. The algorithm first assigned unambiguous correct responses, leaving few ambiguous button presses (presses before 70% coherence of the current scene and after 40% coherence of the following scene or multiple presses occurred on < 5% of trials). Second, ambiguous presses were assigned to an adjacent trial if 1 of the 2 had no response. If both adjacent trials had no response, the press was assigned to the closest trial, unless one was a no-go target, in which case subjects were given the benefit of the doubt that they correctly omitted. Finally, if there were multiple presses that could be assigned to any 1 trial, the fastest response was selected. Slight variations to this algorithm yielded highly similar results, as most button presses showed a 1–1 correspondence with presented images²¹.

Estimation of exGaussian distribution parameters. We estimated exGaussian parameters of RTs (μ , σ , τ) using only correct commission responses (press to city image) for every participant. To estimate parameters, we used maximum likelihood estimation method implemented in SciPy version 1.5.2³⁶. As a comparison, we further estimated the parameters of the Gaussian distribution (μ_{normal} , σ_{normal}) in the same way. We compared which distribution fit the data better by comparing the R -squared values between the actual data and the simulated data obtained using the estimated parameters. Note, we found almost half of participants had negative skewness in their RT distribution, which indicates the long tail for fast RT, in Dataset1 (11,291 of 21,406 participants). Since the assumed exGaussian distribution is not appropriate for negative skewed distributions, we estimated exGaussian parameters (μ , σ , τ) using flipped RT distribution (reverse the positive and negative values of RT) for participants with negative skewness. In this way, we were able to improve the goodness of fit for participants with negative skewness (Supplementary Fig. S1). When we integrate the parameters with the participants with positive skewness, we flipped the value of μ , but used σ and τ to put them on the same scale. Therefore, τ was initially interpreted as the long tail of the RT distribution regardless of which direction in the Dataset1. We note that interpretation of the negative skewness is difficult, as a very slow RT is possibly regarded as a fast RT in the next trial due to the nature of the gradCPT. To support this idea, we found significant difference in mean RT between individuals with positive skewness versus negative skewness (Positive skewness: mean RT = 850 ms; Negative

skewness: mean RT = 888 ms, $t_{21404} = 40.9$, $p < 0.001$, Hedges' $g = 0.56$, two sample t -test). From this result, we can speculate that a very slow RT from individuals with slower RTs on average is likely to be regarded as a fast RT on the next trial (and omission error to current trial). This may explain why individuals with negative skewness had slower mean RT. However, a previous study shows the existence of a type of error called anticipation that indicates pressing a button in anticipation of a stimulus resulting in a very fast RT²⁰. Thus, it is difficult to interpret whether the negative skewness really represents frequent fast RTs in the Dataset1, or rather slow RTs to the previous stimulus. Given this ambiguity, we also conducted the analyses of Dataset1 separately for both positive and negative skewed participants, which revealed mostly consistent results, albeit with some differences. Thus, to avoid this ambiguity, for Dataset2 and Dataset3, as most of participants had positive skewness (81% and 86%, respectively), participants with negative skewness were excluded from the analysis so that the long tail factor can be interpreted as slow RTs unambiguously. We found mean RTs were significantly slower in Dataset1 than those in the Dataset2, which had the same ISI (Dataset1: mean RT = 871 ms; Dataset2: mean RT = 744 ms, $t_{21420} = 7.49$, $p < 0.001$, Hedges' $g = 1.89$, two sample t -test). From this result, we can speculate that this is the reason why there were many participants with negative skewness in Dataset1 with the same logic as above, which could be driven by technical differences in response collection (web vs. lab)². Note that when we included participants with negative skewness in the analysis by estimating exGaussian parameters using flipped RT method described above, the pattern of the results did not change in Dataset2 and Dataset3 (Supplementary Fig. S2).

Calculation of accuracy. In addition to RT related performance, we also calculated accuracy of response to investigate the relationship with exGaussian parameters in Dataset1. Trials in which participants erroneously inhibited a button press to city images were considered omission errors. Trials in which participants erroneously responded to mountain images were considered commission errors. We calculated the Pearson correlation coefficient between exGaussian parameters (μ , σ , τ) and the rate of omission and commission errors, respectively.

fMRI analysis. The following descriptions were exactly the same methodology as in our previous study³².

Preprocessing. A preprocessing of the fMRI data was performed using FMRIprep version 1.3.0³⁷. Preprocessing steps were realignment, coregistration, segmentation of T1-weighted structural images, normalization to Montreal Neurological Institute space. For more details of the pipeline, see <http://fmripred.readthedocs.io/en/latest/workflows.html>.

fMRI signal extraction from brain networks. We used the 400 regions of interest (ROIs) that were defined as 4-mm spheres around the center coordinates³⁸. The blood oxygen level dependent (BOLD) signal time courses were extracted with spatial smoothing using an isotropic Gaussian kernel of 6 mm full-width at half-maximum from these 400 ROIs.

We removed several sources of spurious variance from 400 ROIs' signal time courses by using linear regression with eighteen regression parameters, including six motion parameters, an average signal over the whole brain, five event-related task regressors on the BOLD response time course and six physiological noise regressors. To account for task events (commission error, correct omission, correct commission, and omission error) and trial-to-trial RT, we estimated five BOLD response time courses of each event type by using *hemodynamic_models* function implemented in Nistat (<https://nistat.github.io/>). To account for the physiological confounding, we extracted six physiological noise regressors by applying anatomical CompCor (aCompCor)³⁹. We obtained a mask to exclude signals with a cortical origin by eroding the brain mask and ensuring that it contained subcortical structures only. Six aCompCor components were calculated within the intersection of the subcortical mask and union of the CSF and WM masks calculated in T1-weighted image space after their projection to the native space of functional images in each session. Finally, to account for variance and mean differences across run and participant, we standardized the BOLD signal time course for each ROI (shifted to zero mean and scaled to unit variance) after noise regression. These ROIs were then classified into seven functionally different brain networks that were determined in the previous studies³⁸. A recent study identified two distinct subsystems within the FPN. FPN_A exhibited stronger connectivity with the DMN than the DAN, whereas FPN_B exhibited the opposite pattern⁴⁰. Therefore, we split the FPN into FPN_A and FPN_B using Yeo's 17 network³⁸ and classified into eight functionally different brain networks in the end. For each participant, we then calculated eight time series that represented the activity of these brain networks by averaging BOLD signal time courses in the 400 ROIs corresponding to those brain networks. Note, using the voxel-wise network average time series instead of the ROI-wise network average produced nearly identical patterns of results.

Pairwise maximum entropy model. The pairwise Maximum entropy model (MEM) was applied to the preprocessed BOLD signals as follows in the same manner as that employed in previous studies^{41–46}. To conduct this analysis, we used open Energy Landscape Analysis Toolkit (<https://sites.google.com/site/ezakitakahiro/software>). For each network activity time series, the obtained BOLD signals were binarized with a threshold that was defined as the time-averaged activity of the network activity. We then concatenated BOLD signals from all runs and all participants for each network activity. In this method, the binarized activity σ_i^t at network i and discrete time t represents either active or inactive (+1 or 0). The activity pattern at time t is described by $V^t = [\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t]^T$ where N is the number of the networks. In our case, the number of the networks was eight. The k th brain activity pattern is described by V_k ($k = 1, 2, \dots, 2^N$). The probability distribution of the k th brain activity pattern (V_k) with the largest entropy follows the Boltzmann distribution, when the empirical activation of network i , $\langle \sigma_i \rangle = (1/T) \sum_{t=1}^T \sigma_i^t$, and the empirical pairwise interaction between networks

i and j , $\langle \sigma_i \sigma_j \rangle = (1/T) \sum_{t=1}^T \sigma_i^t \sigma_j^t$ are estimated from the data, where T is the number of volumes⁴⁷. That is, $P(V_k) = e^{E(V_k)} / \sum_{i=1}^{2^N} e^{E(V_i)}$, where $E(V_k)$ is the energy of activity pattern V_k and is given by

$$E(V_k) = -\sum_{i=1}^N h_i \sigma_i(V_k) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N J_{ij} \sigma_i(V_k) \sigma_j(V_k) \quad (1)$$

$\sigma_i(V_k)$ is the binarized activity (+1 or 0) at network i under activity pattern V_k . We estimated h_i and J_{ij} by maximum likelihood estimation to adjust the model-based activation of network i , $\langle \sigma_i \rangle_{model} (= h_i)$ and the model-based pairwise interaction between networks i and j , $\langle \sigma_i \sigma_j \rangle_{model} (= J_{ij})$ toward the empirical activation $\langle \sigma_i \rangle$ and the empirical pairwise interaction $\langle \sigma_i \sigma_j \rangle$, $\langle \sigma_i \rangle_{model} = \sum_{l=1}^{2^N} \sigma_i(V_l) P(V_l)$ and $\langle \sigma_i \sigma_j \rangle_{model} = \sum_{l=1}^{2^N} \sigma_i(V_l) \sigma_j(V_l) P(V_l)$. The likelihood was maximized by a gradient ascent scheme. Please see the previous study for more details how to estimate h_i and J_{ij} ⁴².

We confirmed whether the pairwise MEM accurately fit to the data by calculating Pearson's correlation coefficient between empirical appearance probability and model appearance probability $P(V_k)$. Empirical appearance probability of brain activity pattern V_k is calculated by $(1/T) \sum_{t=1}^T z_k^t$, where \mathbf{z}^t is a K -dimensional binary variable having a 1-of- K representation in which a particular element z_k^t is equal to 1 and all other elements are equal to 0. The values of z_k^t therefore satisfy $z_k^t \in \{0, 1\}$ and $\sum_k z_k^t = 1$, and we see that there are $K (= 2^N)$ possible brain activity patterns for the vector \mathbf{z}^t at time t according to which element is nonzero. In non-technical terms, this confirms whether the model successfully fit the data.

Energy landscape analysis (Definition of brain state). The energy landscape is defined as a graph of brain activity patterns V_k with the corresponding energy $E(V_k)$ as done in the previous studies^{41,42,45,46}. We first exhaustively searched for local energy minima that has the smallest energy value among those of all the N adjacent patterns. Two brain activity patterns are regarded as adjacent if they take the opposite binary activity at just one brain region. Thus, we can discover some local minimum brain activity patterns. Therefore, the number of brain states (local minimum brain states) is determined in a data-driven manner. We then summarized all brain activity patterns into local minimum brain states in the following manner. We first selected a starting brain activity pattern k among the 2^N brain activity patterns. Then, if any of its neighbor patterns has a smaller value of energy than pattern k , we moved to the neighbor pattern with the smallest energy value. Otherwise, we did not move, which implied that pattern k was a local minimum. We repeated this procedure until arriving at a local minimum. The starting pattern k was regarded to belong to the local minimum brain state that was finally reached. We estimated the corresponding local minimum brain state for all brain activity patterns, so that we could estimate the time series of brain states from the time series of brain activity patterns.

exGaussian parameters in each brain state. We observed two dominant brain states during the sustained attention task using the energy landscape analysis³². State1 was characterized by activation of FPN_A, DMN and limbic network, and State2 was characterized as activation of FPN_B, DAN, SAN, SMN and visual network. After defining these brain states, all participants had a brain state transition time series and a RT time series. Thus, RTs were gathered for each brain state, and exGaussian parameters were estimated using the RTs for each brain state in each participant. To account for the hemodynamic response lag, we shifted the time labels of the brain states backwards by 5 s. To statistically investigate the difference in parameters between brain states, we conducted two-tailed paired t -tests and calculated Hedges' g as the effect size^{48,49}. We did not control for multiple comparisons for this analysis, because we confirmed its reproducibility by performing the same analysis on two independent datasets.

Details in each dataset. **Dataset1.** We included 21,406 volunteers between the ages of 10 and 70 years (mean = 30.48, sd = 13.27) in the analyses. Participants were visitors to TestMyBrain.org, a cognitive testing web site where members of the public participants in research studies in exchange for individualized feedback about their performance^{2,50,51}. Data from repeat participation were excluded. Overall, 25,274 participants completed the task, and 21,406 passed quality control measures (missing data or technical problems)². Among technical problems, we chose to exclude testing sessions that exceeded 10% error in the average stimulus presentation time, that is, the time it took for a new image to transition from 0 to 100% opacity. Other participants were excluded for "tune-outs," defined as intervals of 30 s or more without a response. Of the 21,406 participants that were included, there was a nearly equal ratio of males and females (11,540 males, 9827 females, 39 unknown). The study and all analyses were approved by the Committee for the Use of Human Subjects at Harvard University, and written informed consent was obtained from all participants. It was performed in accordance to the relevant guidelines and conforms to the principles of the Declaration of Helsinki. The data used in this study and portions of the methods have been published^{2,51}, but the current analyses and results have not been published elsewhere.

Dataset2. Sixteen participants (6 males, ages 18–34 years, mean age = 24.1 years) performed the gradCPT during fMRI. Subjects completed the three 8-min gradCPT runs. All participants were right-handed, with normal or corrected-to-normal vision and no reported history of major medical illness, head trauma, neurological, or psychiatric disorder. The study and all analyses were approved by the VA Boston Healthcare System IRB, and written informed consent was obtained from all participants. It was performed in accordance to the relevant guidelines and conforms to the principles of the Declaration of Helsinki. The data used in this study and portions of the methods have been previously published^{21,32}, but the current analyses and results have not been published elsewhere.

Descriptive statistics values	Mean	Standard deviation	Skewness	Kurtosis
Positive Skewness group				
μ	0.80	0.06	-0.17	1.02
σ	0.08	0.03	1.55	3.47
τ	0.05	0.02	1.00	7.94
Negative Skewness group				
μ	0.97	0.10	0.43	-0.13
σ	0.10	0.03	1.18	2.30
τ	0.08	0.05	1.55	3.12
All participants				
μ	0.89	0.12	0.51	-0.08
σ	0.09	0.03	1.21	2.27
τ	0.07	0.04	2.04	6.14

Table 1. Descriptive statistics values in exGaussian parameters.

Scanning was performed on a 3 T Siemens MAGNETOM Trio system equipped with a 12-channel head coil, at the VA Boston Neuroimaging Research Center. Functional runs included 248 whole-brain volumes acquired using an echo-planar imaging sequence with the following parameters: repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, flip angle = 90°, acquisition matrix = 64 × 64, in-plane resolution = 3.0 mm², 33 oblique slices, slice thickness = 3, 0.75 mm gap. MPRAGE parameters were as follows: TE = 3.32 ms, TR = 2530 ms, flip angle = 7°, acquisition matrix = 256 × 256, in-plane resolution = 1.0 mm² slice thickness = 1.0 mm.

Dataset3. We attempted to replicate the results using independent Dataset 3. Twenty-nine participants (13 males, ages 21–36 years, mean age = 26.4 years) performed the long ISI gradCPT (1300 ms compared to 800 ms in the Dataset 2) during fMRI. Participants performed the gradCPT, modified here to include thought-probes to investigate mind wandering. Subjects completed the four long ISI gradCPT runs with intermittent thought-probes. Subjects were screened by phone and at an initial visit before the day of neuroimaging, where subjects were also trained on performing the long ISI gradCPT. Exclusion criteria were as follows: current mood, psychotic, or anxiety disorders (excluding simple phobias), ADHD, current use of psychotropic medication, full-scale IQ less than 80, neurological disorders, sensorimotor handicaps, current alcohol or substance abuse/dependence, and claustrophobia. The study and all analyses were approved by the Partners Human Research Institutional Review Board, and written consent was obtained from all participants. It was performed in accordance to the relevant guidelines and conforms to the principles of the Declaration of Helsinki. The data used in this study and portions of the methods have been published^{32,52,53}, but the current analyses and results have not been published elsewhere.

Participants performed the gradCPT including thought-probes. Thought-probes appeared every 44–60 s pseudo-randomly (three possible durations of 44, 52, and 60 s). Upon the thought-probe, a question “To what degree was your focus just on the task or on something else?” was displayed. A continuous scale appeared with far-right and far-left anchors of “only task” and “only else”, respectively. Subjects pressed buttons with their middle and ring fingers to move the scale left and right, and with their thumb to enter their response. Responses were recorded on a graded scale of integers (not visible to the subjects) ranging from 0 (only task) to 100 (only else). The gradCPT immediately resumed after subjects entered their question responses (except for the last thought-probe in the run). This method is the same as the one used in our previous studies^{32,52,53}.

Scanning was performed on a 3 T Siemens CONNECTOM scanner with a custom-made 64-channel phased array head coil, housed at the Athinoula A. Martinos Center for Biomedical Imaging. Functional runs included about 490 whole-brain volumes acquired using a multiband, echo-planar imaging (simultaneous multislice factor of 4) with the following parameters: TR = 1,080 ms, TE = 30 ms, flip angle = 60°, acquisition matrix = 55 × 55, in-plane resolution = 2.0 mm², 68 oblique slices, slice thickness = 2. MPRAGE parameters were as follows: TE = 1.15 ms, TR = 2530 ms, flip angle = 7°, acquisition matrix = 256 × 256, in-plane resolution = 1.0 mm² slice thickness = 1.0 mm.

Results

Relationship between sustained attention performance and distribution parameters in Dataset1.

We first confirmed the exGaussian distribution (or reversed exGaussian) better fit the data than the Gaussian distribution (Positive skewness group, exGaussian: R squared = 0.980, Gaussian: R squared = 0.971, $t_{10114} = 52.9$, $p < 0.001$, Hedges' $g = 0.51$; Negative skewness group, reversed exGaussian: R squared = 0.974, Gaussian: R squared = 0.962, $t_{11290} = 47.1$, $p < 0.001$, Hedges' $g = 0.43$, paired t -test, Supplementary Fig. S1). Note, the goodness of fit (R squared) for Gaussian distribution was still high (R squared > 0.9). Descriptive statistics values of exGaussian parameters are summarized in Table 1. Given the relatively high skew and kurtosis, we report Spearman correlations (a common rule of thumb for normality is ± 2 and 6, respectively, for skew and kurtosis⁵⁴). Furthermore, we assessed the reliability of these parameters by using a split-half method. Namely, we divided into the first 2 min and the second 2 min and fitted exGaussian distribution (or reversed exGaussian) separately to each half, and calculated the Spearman-Brown corrected reliability. We found acceptable to high

	β	t	p
Dependent variable: Omission error			
μ	1.37	22.7	<0.00001
σ	2.78	57.1	<0.00001
τ	3.63	60.2	<0.00001
$R^2=0.398, F(3, 21,402)=4710, p<0.00001$			
Dependent variable: Commission error			
μ	-2.44	-79.1	<0.00001
σ	2.04	81.9	<0.00001
τ	1.63	52.8	<0.00001
$R^2=0.368, F(3, 21,402)=4151, p<0.00001$			

Table 2. Multiple regression predicting each error with exGaussian parameters.

reliability for parameters (All participants, μ : reliability = 0.78; σ : reliability = 0.76; τ : reliability = 0.5). Note that although this version of the task is only 4 min in duration, we confirmed that there were significant performance decrements (accuracy decreased and RT variance increased) over time (Supplementary Fig. S3) indicating this abbreviated version of the task did tax sustained attention.

Next, we investigated relationship between exGaussian parameters and errors by using all participants. We found that μ was moderately positively correlated with the number of omission errors and negatively correlated with the number of commission errors across individuals (Omission errors: Spearman's $\rho = 0.43, p < 0.001$, Supplementary Fig. S4a; Commission errors: Spearman's $\rho = -0.23, p < 0.001$, Supplementary Fig. S4d). This result indicates that the participants who tend to respond slowly (larger μ) have more frequent omission errors and fewer commission errors, and vice versa. In other words, μ could represent an individual strategy factor, which is the same interpretation as average RT and criterion in our previous studies assuming a Gaussian distribution^{2,55}. Next, we found σ was moderately positively correlated with both the number of omission errors and the number of commission errors (Omission error: Spearman's $r = 0.56, p < 0.001$, Supplementary Fig. S4b; Commission error: Spearman's $r = 0.39, p < 0.001$, Supplementary Fig. S4e), consistent with previous interpretation of RT variability as reflecting an ability factor². On the other hand, we found τ was considerably positively correlated with the number of omission errors and weakly positively correlated with the number of commission errors (Omission errors: Spearman's $r = 0.55, p < 0.001$, Supplementary Fig. S4c; Commission errors: Spearman's $r = 0.17, p < 0.001$, Supplementary Fig. S4f). Furthermore, a multiple regression analysis with each error as the dependent variable, yielded significant coefficients for every parameter differently (see Table 2). We found that the number of omission errors was best explained by τ and the number of commission errors was best explained by σ . These results indicate that fitting exGaussian could separate the strategy factor (μ), the variability factor (σ), and the long tail factor (τ), because all parameters have different relationship with sustained attention performance. We next investigated the relationships in negative skewness group and positive skewness group separately. The patterns of the result were similar except the relationship between μ and the number of omission error was absent in positive skewness group (Fig. 2). Given these differences, we focused on the un-ambiguous positive skew group in the subsequent datasets, but report the results from the full sample for completeness in the Supplementary Information.

Difference in exGaussian distribution parameters between brain states in Dataset2. As we have shown in our previous studies³², we observed two dominant brain states during gradCPT using an energy landscape analysis (Fig. 3b). State1 was characterized by higher activity in FPN_A, DMN and limbic network, and was behaviorally optimal (high accuracy), while State2 was characterized by higher activity in FPN_B, DAN, SN, SMN and visual network and was behaviorally suboptimal (low accuracy) (Supplementary Fig. S5a). This indicates that fluctuation of brain activity during gradCPT can be described as dynamic transitions between the two brain states represented by State1 and State2. In Dataset2, three of 16 participants had negative skewness and were excluded from the analysis. We first confirmed the exGaussian distribution better fit the data than the Gaussian distribution ($t_{12} = 3.64, p < 0.005$, Hedges' $g = 0.78$, paired t -test, Fig. 3a,c). We then compared the exGaussian parameters between two brain states. We found only σ was significantly larger in State2 than State1 (μ : $t_{12} = 1.35, p = 0.20$, Hedges' $g = 0.083$; σ : $t_{12} = 2.29, p < 0.05$, Hedges' $g = 0.31$; τ : $t_{12} = 0.32, p = 0.76, g = 0.048$, paired t -test, Fig. 3b,d). This result indicates that the difference in RT variance between brain states was not due to difference in the long tail factor, but due to the difference in the variability factor. Note that when we estimated exGaussian parameters using flipped RT distribution for participants with negative skewness and included in analysis, the pattern of results did not change (Supplementary Fig. S2a).

Replication using an independent dataset. We observed the same two dominant brain states in Dataset3 (Fig. 4b). State1 was characterized by higher activity in FPN_A, DMN and limbic network, and was behaviorally optimal (high accuracy), while State2 was characterized by higher activity in FPN_B, DAN, SN, SMN and visual network and was behaviorally suboptimal (low accuracy) (Supplementary Fig. S5b). In Dataset3, four of

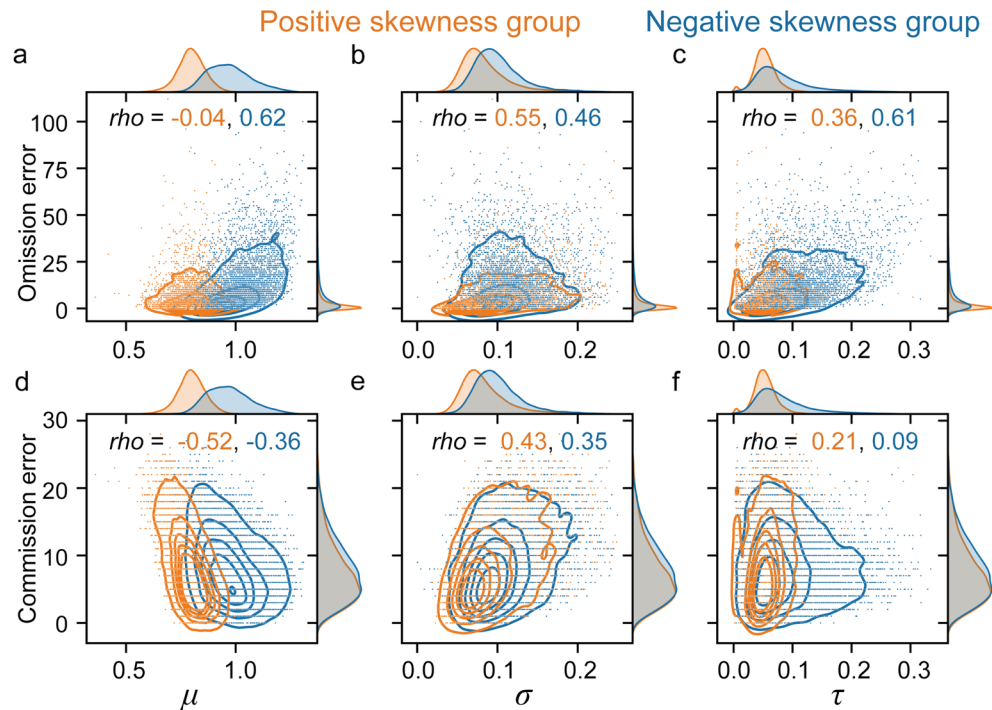


Figure 2. Relationship between exGaussian distribution parameters and sustained attention performance in each skewness group. (a) Scatter plot and histograms of μ and the number of omission errors. (b) Scatter plot and histograms of σ and the number of omission errors. (c) Scatter plot and histograms of τ and the number of omission errors. (d) Scatter plot and histograms of μ and the number of commission errors. (e) Scatter plot and histograms of σ and the number of commission errors. (f) Scatter plot and histograms of τ and the number of commission errors. Solid line indicates a kernel density estimate, which is a method for visualizing the distribution of observations. Spearman's correlation coefficients values were shown in each panel. Orange color indicates positive skewness group and blue color indicates negative skewness group.

29 participants had negative skewness and were excluded from analyses. We first confirmed that the exGaussian distribution better fit the data than the Gaussian distribution ($t_{24}=6.52$, $p<0.001$, Hedges' $g=0.94$, paired t -test, Fig. 4a,c). We then compared exGaussian parameters between two brain states. We found only σ was significantly larger in State2 than State1 (μ : $t_{24}=1.78$, $p=0.089$, Hedges' $g=0.099$; σ : $t_{24}=6.29$, $p<0.001$, Hedges' $g=0.45$; τ : $t_{24}=0.46$, $p=0.65$, Hedges' $g=0.069$, paired t -test, Fig. 4b,d). Again, σ was higher in the optimal vs suboptimal brain state, which coincided with worse accuracy (Supplementary Fig. S5b). This result indicates that we successfully replicated our result in the independent dataset. Note that when we included estimated participants with negative skewness by estimating exGaussian parameters using flipped RT distribution, the pattern of results did not change (Supplementary Fig. S2b).

We further investigated the relationship between exGaussian parameters and mind wandering score across participants. We found a significant Spearman's correlation coefficients between σ and mind wandering score (μ : $rho=0.17$, $p=0.38$; σ : $rho=0.56$, $p=0.0018$; τ : $rho=0.37$, $p=0.0502$, Supplementary Fig. S6). That is, the participants who tend to mind wandering had larger σ . Although we did not find significant correlation between τ and mind wandering score, statistical power may have been insufficient to find the relationship ($rho=0.37$).

Discussion

In the first part of this study, we aimed to disambiguate the ways in which reaction times are behavioral indicators of sustained attention. To do this, we verified the utility of exGaussian fitting to RT distributions in the gradCPT, and investigated the relationship between exGaussian parameters (μ , σ , and τ) and sustained attention performance (the number of omission error and commission error) using a large gradCPT dataset which included over 20,000 participants. We found that all exGaussian parameters were uniquely related to different aspects of sustained attention performance across individuals, with the variance factor most consistently related to all measures of accuracy. In the second part of this study, we tested the hypothesis that this variance factor represents different brain states (large variance hypothesis) as identified by an fMRI energy landscape analysis, vs. the alternative hypothesis that the long tail factor represents these different brain states (long tail hypothesis). Our results showed a significant difference in the variance factor (σ) between brain states, such that one brain state reflects optimal attention (lower variance and higher accuracy) and the other reflects suboptimal attention (higher variance and lower accuracy). We further replicated these results in a second independent dataset.

We found that μ , or the mean of the normal component of the reaction time, had moderate positive correlation with omission error and negative correlation with commission error. This result is consistent with previous

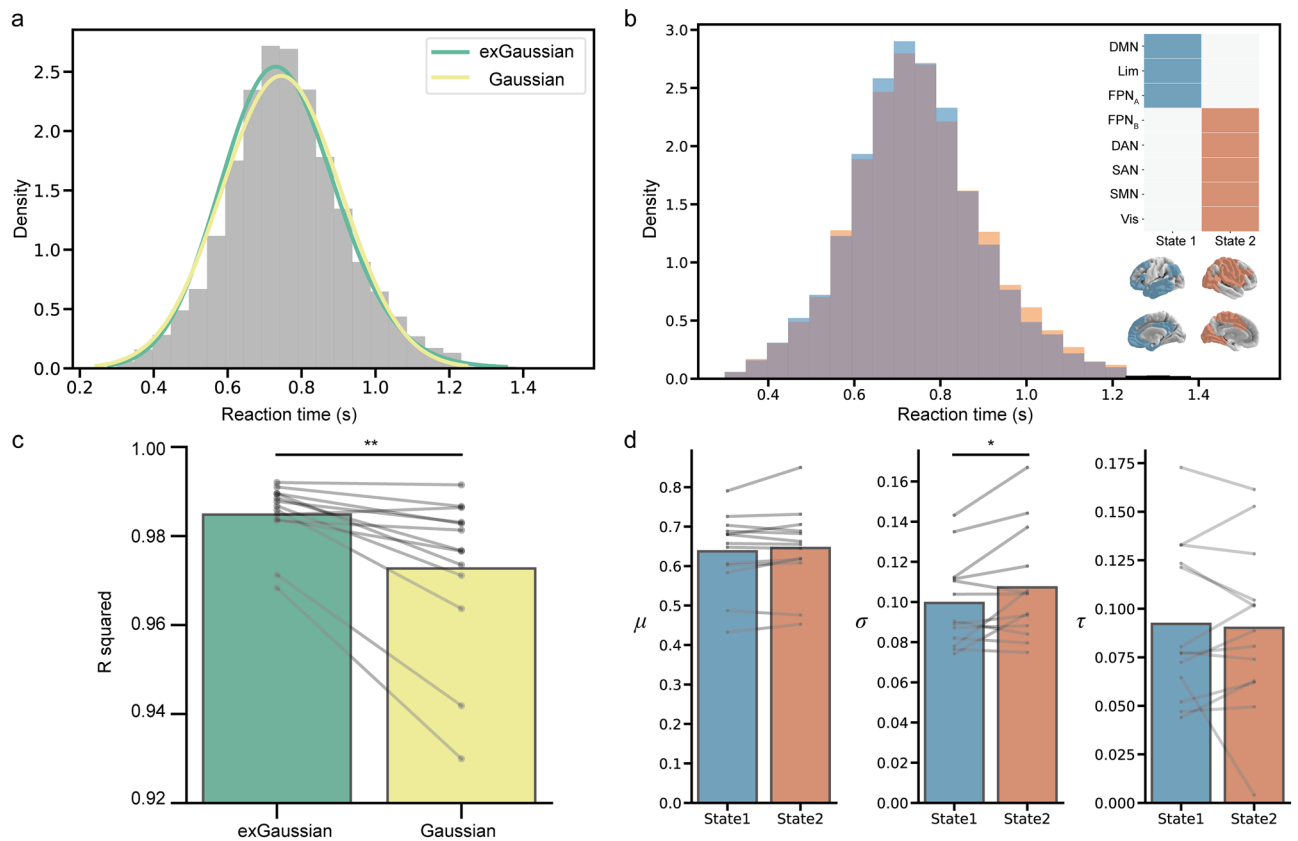


Figure 3. Results summary in Dataset2. **(a)** RT histogram and fitting result. **(b)** RT histograms in each brain state. Individual state is represented by an activity pattern in which each brain region is active (blue and red) or inactive (white). **(c)** R squared values for exGaussian and Gaussian distributions for each individual. **(d)** exGaussian parameters differences between brain states for each individual. * $p < 0.05$, ** $p < 0.005$. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual.

reports that fast RTs are associated with commission error and slow RTs are associated with omission error²⁰. That is, it is thought that μ represents either a strategy to reduce omission errors by speeding up the average RT or a strategy to reduce commission errors by slowing down the average RT^{2,29,56}. This strategy factor, related to speed of reaction time, is also associated with a response bias to press (faster) or withhold (slower)². We also found that σ , or the variance of the normal component of the reaction time, had moderate positive correlation with both omission error and commission error. We also found that σ was related to mind wandering. These results indicate that the variance factor is related to errors in general, regardless of the type of error. On the other hand, we found τ , or the exponential tail of the reaction time distribution, had moderate to large positive correlation with omission error but a smaller correlation with commission error. This result suggests that τ was strongly associated with occasional, more catastrophic attention lapses, or task disengagement^{29,56}. These results indicate that fitting exGaussian could separate the strategy factor (μ), the variability factor (σ), and the long tail factor (τ), because all parameters have different relationships with sustained attention performance (the number of omission and commission errors). However, we note that interpretation of τ is difficult as we described in the “Methods” Section (“Estimation of exGaussian distribution parameters”), and therefore the interpretation should be treated with caution. Namely, in the gradCPT, this tail may manifest as both slow RTs (positive skew) or fast RT (negative skew), as slow responses to a given trial may be misassigned as fast responses to the next trial, given the gradual fading and some unavoidable ambiguity of a minority of responses (see “Methods”; Esterman et al. 2013)²¹. Similarly, given the response deadline implicit in this task, the long tail parameter may have been underestimated, and the reliability of this factor was far lower than the other RT factors, potentially reducing the ability to detect meaningful individual differences.

In our previous study³², we found across 4 independent datasets, that two dominant brain states, which were related to optimal and suboptimal attention, occurred during sustained attention. State1 was characterized by higher activity in FPN_A, DMN and limbic network, and was behaviorally optimal, while State2 was characterized by higher activity in FPN_B, DAN, SN, SMN and visual network and was behaviorally suboptimal. RT variance was lower and accuracy was higher during State1 than during State2, suggesting State1 was optimal and State2 was suboptimal. However, we did not previously analyze differences in RTs while accounting for their non-normality.

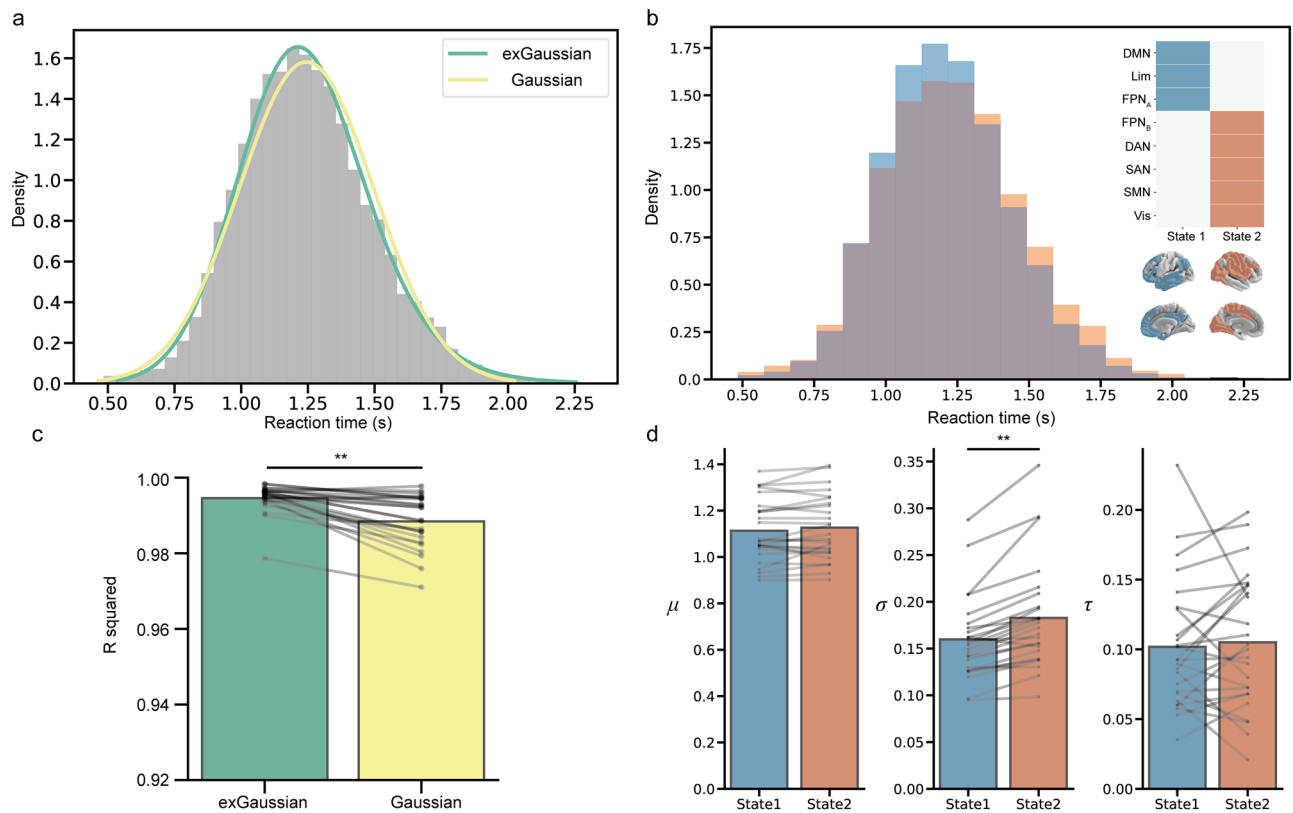


Figure 4. Results summary in Dataset3. **(a)** RT histogram and fitting result. **(b)** RT histograms in each brain state. Individual state is represented by an activity pattern in which each brain region is active (blue and red) or inactive (white). **(c)** R squared values for exGaussian and Gaussian distributions for each individual. **(d)** exGaussian parameters differences between brain states for each individual. $**p < 1.0 \times 10^{-5}$. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual. DMN: default mode network; Lim: limbic; FPN: frontoparietal network; DAN: dorsal attention network; SAN: salience network; SMN: somatomotor network; Vis: visual.

In this study, we examined how RT distribution differed across two dominant brain states by comparing these exGaussian parameters characterized behaviorally in Dataset1. When doing so, we consistently found the variability factor (σ) was significantly different between brain states in two independent datasets. Note, although we did not find a significant difference in the long tail factor (τ) between brain states, it is possible that τ was not different due to the nature of the experimental task. That is, since the gradCPT has a short ISI (800 or 1300 ms), extremely long RTs were not possible. This may also have contributed to the relatively lower reliability of this factor. Despite this limitation, these results suggest that both slow and fast RTs are induced by same suboptimal brain state (State2) characterized as activation of DAN, SAN, SMN, FPN_B and visual networks.

There are a number of potential interpretations of the variance factor. A previous study suggested that variability may represent exploration from the viewpoint of exploring (variable) and exploiting (stable) trade-offs³³. For example, participants with large variability exhibit faster learning than participants with small variability⁵⁷. We speculate that State1 may be the exploitative state and State2 may be the explorative state, with regard to response strategy or approach to the task. For example, when accuracy is higher, participants may continue to use the same approach, leading to more consistent RTs, whereas when performance is worse, participants may explore different approaches, leading to more variable RTs (both faster and slow). Alternatively, it could be that greater RT variability reflects greater noise in the perceptual, selection, or response system. For example, Rothlein et al. (2018) found that during more variable periods of task performance, representational fidelity of stimulus representations was weaker in the visual cortices⁵⁸. A further related possibility is that suboptimal states represent less automatic processing of the stimuli alongside higher perceptual load^{21,55}, which may lead to greater complexity of behavior reflected in higher variability.

While variability during sustained attention is likely to fluctuate due to intrinsic (task-unrelated) factors, there are a number of ways in which extrinsic factors can impact response variability. Our previous studies show that variability is reduced in conditions where participants are rewarded based on their performance^{55,59,60}. Furthermore, motor variability is actively regulated by recent reward history, with variability increasing when

performance is poor^{33,61}. On the other hand, our previous study found reactive DAN/SN activation in response to errors^{60,62}. This suggests that the more variable and suboptimal brain state (State2) also may be partly driven by response to errors. Together, this kind of reward (error)-dependent regulation of variability can be interpreted as an exploitation-exploration dilemma. In other words, while the reward is being obtained, variance is reduced and the current action (or strategy) is continued; when the reward is not obtained (or the participant makes an error), variance increases and another action is taken to change the current action. There is no consensus whether or how variability is regulated in the brain^{63,64}. Further study is needed to investigate when brain state transition occurs, as well how variability changes in relationship to extrinsic factors.

As a limitation of our current study, although we did not find a significant difference in the long tail factor (τ) between brain states, our results do not rule out the possibility that fast and slow RTs are induced from different brain mechanisms (fast RTs represent mindless and automatic processing, and slow RTs represent disengaged/inefficient processing). This is because it is possible that τ was not different due to the nature of the gradCPT as we discussed above. As such, these results may not generalize beyond the gradCPT, or other CPTs with short response deadlines. The similarities and differences in cognitive mechanisms between gradCPT and other sustained attention task such as PVT, which do not have a response deadline, are important and interesting research topics. Nonetheless, given the ubiquity of the gradCPT and similar task in the field^{2,4,14}, the current results are relevant to the field. Critically, our results provide evidence of the existence of a stable RT brain state and a variable RT brain state independent of a long tail during this type of sustained attention task. Another limitation is that our results do not address any causal relationship between brain state and behavior. Some evidence suggests, from our previous studies, that activation of DAN and salience network, which are included in suboptimal State2, occur in response to omission and commission errors^{26,60}. Therefore, the State2-related networks may be driven by reactivity to errors or increased variability. Similarly, DMN activity is associated with the optimal state (State1) and low variability, but also higher degrees of mind wandering during gradCPT³⁵, suggesting a complex role of DMN in sustained attention. Therefore, further study is needed to clarify the causal relationships between network activity and sustained attention. This may be possible by using real time presentation of stimuli according to brain state⁶⁵. Additionally, event-related TMS of these networks could further help infer causality^{66,67}.

In summary, we revealed all parameters of the exGaussian were uniquely related to sustained attention performance based on data from a large sample. μ is the strategy factor, σ is the variability factor, and τ is the long tail factor, respectively. Our results showed the utility of the exGaussian distribution for relating RTs to sustained attention. We further revealed that the variance factor is significantly different between two dominant brain states during a sustained attention task. This result suggests that the suboptimal brain State2, relative to the optimal brain State1, is not characterized by slower RTs, but rather by increased variance of RTs (and reduced accuracy). We believe that our findings better identify and refine a behavioral marker of optimal sustained attention based on brain states, and represent an important theoretical and methodological finding for future sustained attention research.

Data availability

Analysis code and summary data required to reproduce all figures in our manuscript is publically available (https://github.com/Ayumu722/ScientificReports_exGaussian). However, a portion of the raw data (e.g. brain images) are owned by the United States Department of Veterans Affairs and are available only upon request from the United States Department of Veterans Affairs. The Department of Veterans Affairs will make this data publicly available and requests for the data can be made by interested individuals by filing a Freedom of Information Act request to the Privacy Officer at VA Boston Healthcare System (vhabhsFOIAofficers@va.gov) or the FOIA Intake Center (see http://www.oprm.va.gov/foia/foia_howTo.aspx for more details).

Received: 6 April 2021; Accepted: 7 July 2021

Published online: 21 July 2021

References

- Edkins, G. D. & Pollock, C. M. The influence of sustained attention on railway accidents. *Accid. Anal. Prev.* **29**, 533–539. [https://doi.org/10.1016/S0001-4575\(97\)00033-X](https://doi.org/10.1016/S0001-4575(97)00033-X) (1997).
- Fortenbaugh, F. C. *et al.* Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychol. Sci.* **26**, 1497–1510. <https://doi.org/10.1177/0956797615594896> (2015).
- Isbell, E., Calkins, S. D., Swingler, M. M. & Leerkes, E. M. Attentional fluctuations in preschoolers: Direct and indirect relations with task accuracy, academic readiness, and school performance. *J. Exp. Child. Psychol.* **167**, 388–403. <https://doi.org/10.1016/j.jecp.2017.11.013> (2018).
- Madore, K. P. *et al.* Memory failure predicted by attention lapsing and media multitasking. *Nature* **587**, 87–91. <https://doi.org/10.1038/s41586-020-2870-z> (2020).
- Robertson, I. Attention and error. *Psychologist* **16**, 476 (2003).
- Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P. & Tannock, R. Characterizing cognition in ADHD: Beyond executive dysfunction. *Trends Cogn. Sci.* **10**, 117–123. <https://doi.org/10.1016/j.tics.2006.01.011> (2006).
- Castellanos, F. X. & Tannock, R. Neuroscience of attention-deficit/hyperactivity disorder: The search for endophenotypes. *Nat. Rev. Neurosci.* **3**, 617–628. <https://doi.org/10.1038/nrn896> (2002).
- Klein, C., Wendling, K., Huettner, P., Ruder, H. & Peper, M. Intra-subject variability in attention-deficit hyperactivity disorder. *Biol. Psychiatry* **60**, 1088–1097. <https://doi.org/10.1016/j.biopsych.2006.04.003> (2006).
- Hammar, A. & Ardal, G. Cognitive functioning in major depression: A summary. *Front. Hum. Neurosci.* **3**, 26. <https://doi.org/10.3389/neuro.09.026.2009> (2009).
- Cornblatt, B. A. & Keilp, J. G. Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophr. Bull.* **20**, 31–46. <https://doi.org/10.1093/schbul/20.1.31> (1994).
- Martínez-Arán, A. *et al.* Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. *Am. J. Psychiatry* **161**, 262–270. <https://doi.org/10.1176/appi.ajp.161.2.262> (2004).

12. DeGutis, J. *et al.* Posttraumatic psychological symptoms are associated with reduced inhibitory control, not general executive dysfunction. *J. Int. Neuropsychol. Soc.* **21**, 342–352. <https://doi.org/10.1017/S1355617715000235> (2015).
13. Esterman, M. *et al.* Trauma-related psychiatric and behavioral conditions are uniquely associated with sustained attention dysfunction. *Neuropsychology* **33**, 711–724. <https://doi.org/10.1037/neu0000525> (2019).
14. Fortenbaugh, F. C., DeGutis, J. & Esterman, M. Recent theoretical, neural, and clinical advances in sustained attention research. *Ann. N. Y. Acad. Sci.* **1396**, 70–91. <https://doi.org/10.1111/nyas.13318> (2017).
15. Uddo, M., Vasterling, J. J., Brailey, K. & Sutker, P. B. Memory and attention in combat-related post-traumatic stress disorder (PTSD). *J. Psychopathol. Behav. Assess.* **15**, 43–52. <https://doi.org/10.1007/BF00964322> (1993).
16. Fortenbaugh, F. C. *et al.* Interpersonal early-life trauma alters amygdala connectivity and sustained attention performance. *Brain Behav.* **7**, e00684. <https://doi.org/10.1002/brb3.684> (2017).
17. Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. & Yiend, J. Oops!: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* **35**, 747–758. [https://doi.org/10.1016/S0028-3932\(97\)00015-8](https://doi.org/10.1016/S0028-3932(97)00015-8) (1997).
18. Mackworth, N. H. The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* **1**, 6–21. <https://doi.org/10.1080/17470214808416738> (1948).
19. Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. Jr. & Beck, L. H. A continuous performance test of brain damage. *J. Consult. Psychol.* **20**, 343–350. <https://doi.org/10.1037/h0043220> (1956).
20. Allan Cheyne, J., Solman, G. J., Carriere, J. S. & Smilek, D. Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition* **111**, 98–113. <https://doi.org/10.1016/j.cognition.2008.12.009> (2009).
21. Esterman, M., Noonan, S. K., Rosenberg, M. & Degutis, J. In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* **23**, 2712–2723. <https://doi.org/10.1093/cercor/bhs261> (2013).
22. Osmon, D. C., Kazakov, D., Santos, O. & Kassel, M. T. Non-Gaussian distributional analyses of reaction times (RT): Improvements that increase efficacy of RT tasks for describing cognitive processes. *Neuropsychol. Rev.* **28**, 359–376. <https://doi.org/10.1007/s11065-018-9382-8> (2018).
23. Smallwood, J. *et al.* Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Conscious Cogn.* **13**, 657–690. <https://doi.org/10.1016/j.concog.2004.06.003> (2004).
24. MacDonald, S. W., Li, S. C. & Backman, L. Neural underpinnings of within-person variability in cognitive functioning. *Psychol. Aging* **24**, 792–808. <https://doi.org/10.1037/a0017798> (2009).
25. MacDonald, S. W., Nyberg, L. & Backman, L. Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends Neurosci.* **29**, 474–480. <https://doi.org/10.1016/j.tins.2006.06.011> (2006).
26. Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J. & Esterman, M. Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension. *Neuroimage* **171**, 148–164. <https://doi.org/10.1016/j.neuroimage.2018.01.002> (2018).
27. Hohle, R. H. Inferred components of reaction times as functions of foreperiod duration. *J. Exp. Psychol.* **69**, 382–386. <https://doi.org/10.1037/h0021740> (1965).
28. Hohle, R. H. Component process latencies in reaction times of children and adults. *Adv. Child Dev. Behav.* **3**, 225–261. [https://doi.org/10.1016/S0065-2407\(08\)60456-5](https://doi.org/10.1016/S0065-2407(08)60456-5) (1967).
29. Gu, S. L., Gau, S. S., Tzang, S. W. & Hsu, W. Y. The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Res. Dev. Disabil.* **34**, 3709–3719. <https://doi.org/10.1016/j.ridd.2013.07.025> (2013).
30. Lin, H. Y. *et al.* Neural substrates of behavioral variability in attention deficit hyperactivity disorder: Based on ex-Gaussian reaction time distribution and diffusion spectrum imaging tractography. *Psychol. Med.* **44**, 1751–1764. <https://doi.org/10.1017/S0033291713001955> (2014).
31. Nir, Y. *et al.* Selective neuronal lapses precede human cognitive lapses following sleep deprivation. *Nat. Med.* **23**, 1474–1480. <https://doi.org/10.1038/nm.4433> (2017).
32. Yamashita, A., Rothlein, D., Kucyi, A., Valera, E. M. & Esterman, M. Brain state-based detection of attentional fluctuations and their modulation. *Neuroimage* **236**, 118072. <https://doi.org/10.1016/j.neuroimage.2021.118072> (2021).
33. Dhawale, A. K., Miyamoto, Y. R., Smith, M. A. & Olveczky, B. P. Adaptive regulation of motor variability. *Curr. Biol.* **29**, 3551–3562. <https://doi.org/10.1016/j.cub.2019.08.052> (2019).
34. Leth-Steensen, C., Elbaz, Z. K. & Douglas, V. I. Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Physiol. (Oxf)* **104**, 167–190. [https://doi.org/10.1016/S0001-6918\(00\)00019-6](https://doi.org/10.1016/S0001-6918(00)00019-6) (2000).
35. Kucyi, A., Esterman, M., Riley, C. S. & Valera, E. M. Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proc. Natl. Acad. Sci. USA* **113**, 13899–13904. <https://doi.org/10.1073/pnas.1611743113> (2016).
36. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2> (2020).
37. Esteban, O. *et al.* fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116. <https://doi.org/10.1038/s41592-018-0235-4> (2019).
38. Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114. <https://doi.org/10.1093/cercor/bhx179> (2018).
39. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042> (2007).
40. Dixon, M. L. *et al.* Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proc. Natl. Acad. Sci. USA* **115**, E1598–E1607. <https://doi.org/10.1073/pnas.1715766115> (2018).
41. Ezaki, T., Sakaki, M., Watanabe, T. & Masuda, N. Age-related changes in the ease of dynamical transitions in human brain activity. *Hum. Brain Mapp.* **39**, 2673–2688. <https://doi.org/10.1002/hbm.24033> (2018).
42. Ezaki, T., Watanabe, T., Ohzeki, M. & Masuda, N. Energy landscape analysis of neuroimaging data. *Philos. Trans. A* <https://doi.org/10.1098/rsta.2016.0287> (2017).
43. Watanabe, T. *et al.* A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nat. Commun.* **4**, 1370. <https://doi.org/10.1038/ncomms2388> (2013).
44. Watanabe, T. *et al.* Energy landscapes of resting-state brain networks. *Front. Neuroinform.* **8**, 2014.00012 (2014).
45. Watanabe, T., Masuda, N., Megumi, F., Kanai, R. & Rees, G. Energy landscape and dynamics of brain activity during human bistable perception. *Nat. Commun.* **5**, 4765. <https://doi.org/10.1038/ncomms5765> (2014).
46. Watanabe, T. & Rees, G. Brain network dynamics in high-functioning individuals with autism. *Nat. Commun.* **8**, 16048. <https://doi.org/10.1038/ncomms16048> (2017).
47. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630. <https://doi.org/10.1103/PhysRev.106.620> (1957).
48. Bonett, D. G. Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychol. Methods* **14**, 225–238. <https://doi.org/10.1037/a0016619> (2009).
49. Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.* **6**, 107–128. <https://doi.org/10.3102/10769986006002107> (1981).

50. Germine, L. *et al.* Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychon. Bull. Rev.* **19**, 847–857. <https://doi.org/10.3758/s13423-012-0296-9> (2012).
51. Ritz, H., DeGutis, J., Frank, M. J., Esterman, M. & Shenhav, A. An evidence accumulation model of motivational and developmental influences over sustained attention. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*. Toronto, CA.
52. Kucyi, A. *et al.* Prediction of stimulus-independent and task-unrelated thought from functional brain networks. *Nat. Commun.* **12**, 1793. <https://doi.org/10.1038/s41467-021-22027-0> (2021).
53. Zuberer, A. *et al.* Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *Neuroimage* **229**, 117610. <https://doi.org/10.1016/j.neuroimage.2020.117610> (2021).
54. Curran, P. J., West, S. G. & Finch, J. F. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychol. Methods* **1**, 16 (1996).
55. Esterman, M., Reagan, A., Liu, G., Turner, C. & DeGutis, J. Reward reveals dissociable aspects of sustained attention. *J. Exp. Psychol. Gen.* **143**, 2287–2295. <https://doi.org/10.1037/xge0000019> (2014).
56. Ballard, J. C. Assessing attention: Comparison of response-inhibition and traditional continuous performance tests. *J. Clin. Exp. Neuropsychol.* **23**, 331–350. <https://doi.org/10.1076/jcen.23.3.331.1188> (2001).
57. Wu, H. G., Miyamoto, Y. R., Gonzalez Castro, L. N., Olveczky, B. P. & Smith, M. A. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* **17**, 312–321. <https://doi.org/10.1038/nn.3616> (2014).
58. Rothlein, D., DeGutis, J. & Esterman, M. Attentional fluctuations influence the neural fidelity and connectivity of stimulus representations. *J. Cogn. Neurosci.* **30**, 1209–1228. https://doi.org/10.1162/jocn_a_01306 (2018).
59. Esterman, M. *et al.* Anticipation of monetary reward can attenuate the vigilance decrement. *PLoS ONE* **11**, e0159741. <https://doi.org/10.1371/journal.pone.0159741> (2016).
60. Esterman, M., Poole, V., Liu, G. & DeGutis, J. Modulating reward induces differential neurocognitive approaches to sustained attention. *Cereb. Cortex* **27**, 4022–4032. <https://doi.org/10.1093/cercor/bhw214> (2017).
61. Pekny, S. E., Izawa, J. & Shadmehr, R. Reward-dependent modulation of movement variability. *J. Neurosci.* **35**, 4015–4024. <https://doi.org/10.1523/JNEUROSCI.3244-14.2015> (2015).
62. Kucyi, A. *et al.* Electrophysiological dynamics of antagonistic brain networks reflect attentional fluctuations. *Nat. Commun.* **11**, 325. <https://doi.org/10.1038/s41467-019-14166-2> (2020).
63. Dhawale, A. K., Smith, M. A. & Olveczky, B. P. The role of variability in motor learning. *Annu. Rev. Neurosci.* **40**, 479–498. <https://doi.org/10.1146/annurev-neuro-072116-031548> (2017).
64. Sternad, D. It's not (only) the mean that matters: Variability, Noise and exploration in skill learning. *Curr. Opin. Behav. Sci.* **20**, 183–195. <https://doi.org/10.1016/j.cobeha.2018.01.004> (2018).
65. Hinds, O. *et al.* Roles of default-mode network and supplementary motor area in human vigilance performance: Evidence from real-time fMRI. *J. Neurophysiol.* **109**, 1250–1258. <https://doi.org/10.1152/jn.00533.2011> (2013).
66. Esterman, M. *et al.* Frontal eye field involvement in sustaining visual attention: Evidence from transcranial magnetic stimulation. *Neuroimage* **111**, 542–548. <https://doi.org/10.1016/j.neuroimage.2015.01.044> (2015).
67. Esterman, M. *et al.* Network-targeted cerebellar transcranial magnetic stimulation improves attentional control. *Neuroimage* **156**, 190–198. <https://doi.org/10.1016/j.neuroimage.2017.05.011> (2017).

Acknowledgements

This research was supported by Overseas Research Fellow in Japan Society for the Promotion of Science to AY and a Merit Review Award from the Department of Veterans Affairs Clinical Sciences Research and Development (I01CX001653) to ME.

Author contributions

M.E., E.V., L.G., J.W., J.D. and A.K. designed the study and recruited participants for the study, collected their clinical and imaging data. A.Y. performed data preprocessing and analysis; A.Y., D.R., and M.E. primarily wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-94161-0>.

Correspondence and requests for materials should be addressed to A.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021