communications psychology

Article

A Nature Portfolio journal

https://doi.org/10.1038/s44271-026-00424-9

Inferring mind wandering from perceptual decision making

Check for updates

Cathy Zhang ✉ & Wouter Kool

People need to sustain focused attention to achieve goals. Yet, attention often lapses, as minds wander towards task-unrelated thoughts. The conventional way to study such shifts in attention is through thought probes that explicitly ask if thoughts are task-related. However, probes are rare and interrupt behavior. Other methods to measure mind wandering assume a 50/50 split in time spent on-task vs off-task. We address these issues with a framework to infer mind wandering (MW) using computational modeling. We use a random dot motion task with varying evidence, but with a strong bias inducing a repetitive response requirement. Occasional thought probes were used for validation. When participants ($N = 93$) reported being off-task, accuracy was higher and reaction time (RT) was lower, suggesting less stimulus processing and more reliance on bias. To classify internal states for individual trials from performance, we fit a Hidden Markov Model with Generalized Linear Models (GLM-HMM) for each state to responses. A two-state GLM-HMM predicted lower RTs on off-task trials, revealed an increase in mind wandering across the task, and aligned with self-reported focus. This shows that temporal variation in attentional states can be measured on a trial-to-trial basis without thought probes, paving the way for future MW research.

To achieve goals, humans need to exert sustained attentional control. Yet, their minds often drift toward thoughts unrelated to ongoing events[1]. In the language of cognitive psychology, such "mind wandering" is typically defined as a failure to maintain sustained attention on the current task accompanied by a shift in attention towards task-unrelated thoughts (TUTs)[2–7].

Beyond minor disturbances in everyday life (such as missing your bus stop), mind wandering also carries serious consequences. Mind wandering increases negative mood[1], leads to a higher likelihood of traffic accidents[8–10] and impaired academic performance[11–13], and has been related to clinical disorders such as depression[14,15] and attention-deficit hyperactivity disorder[16,17]. Therefore, the study of mind wandering, and its behavioral characteristics[18], temporal dynamics[19,20], and neural correlates[21], has been of widespread scientific interest.

In this pursuit, cognitive psychologists have developed several experimental techniques to capture mind wandering in laboratory settings. However, these tools carry several notable limitations. As detailed below, these techniques either rely on explicit self-report or do not capture between-subject variability in the rate of mind wandering. In this paper, we first review these techniques and then propose a task that enables inferring mind wandering and its temporal dynamics from behavior using computational modeling.

The gold-standard paradigm for studying mind wandering is the Sustained Attention to Response Task (SART)[22]. In this "go/no-go" task, participants make the same response to "go" stimuli, which appear in the majority of the trials, and withhold their response to a "no-go" stimulus, which occurs with a low probability[2,23,24]. The repetitive response requirements of this task quickly become automatic, leading to reduced sustained attention[22,25,26]. Indeed, participants in the SART often fail to withhold their responses in the rare no-go trials. Reduced response times (RT) preceding such errors further suggest that this is caused by lapses in attentional control[22].

Most mind-wandering studies, however, assess the occurrence of mind wandering during the task using *thought probes*[27–30]. These interrupt the task and require participants to explicitly report whether they were focused on task-related or TUTs[4,31,32]. A large body of work using variants of the SART shows that these self-reports capture fluctuations in mental states[27,33,34]. For example, self-reported mind wandering is associated with shorter RTs and higher errors on rare trials in the SART[2,6,25,35,36]. Self-reports of mind wandering have also revealed the existence of two distinct modes of mind wandering (intentional vs. unintentional[37,38]). At the same time, this approach has several well-acknowledged limitations.

First, thought probes occur infrequently and therefore only provide insight into mind wandering for a few time points in the task. Such sparse

Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA. ✉e-mail: cathyz@wustl.edu

sampling is undesirable, especially given that mind wandering likely persists for more than a single trial. By infrequently sampling attentional states, researchers likely fail to capture dynamic changes in attention allocation across time[39]. Second, thought probes can interrupt the natural flow of attentional states. Previous work suggests that probes reduce the occurrence of mind wandering by redirecting participants' awareness to evaluating their thoughts[40]. Third, the thought probe method depends on participants' honesty[41] and metacognitive awareness[42,43], which biases mind-wandering reports. Finally, thought probes come in a wide variety of formats[25,27]. It has been shown that the format of the thought probes affects reports[34,44], with more reported mind wandering when people are asked whether their mind was on "something other than the task" compared to "on the task" [34]. This limits the ability to compare findings across studies[32].

Because of these limitations, the field needs methods that measure mind wandering from behavior alone. One influential idea has been to use variability in RTs as one such marker of fluctuations in attention[45]. Specifically, increased RT variability predicts higher error rates[20,45,46] and more self-reported mind wandering[45–49]. Even though this framework can identify trials with reduced attentional focus, its analytic approach makes limiting assumptions. First, trial-to-trial fluctuations in RT variability are calculated using a moving window to smooth across trials, requiring researchers to pick a timescale at which attention fluctuations occur[45]. Thus, this approach is not sensitive to individual differences in the frequency of attentional state switches. Second, trials are labeled as "on-" and "off-task" by comparing each trial's RT variability to the participant's median. Therefore, this approach results in an equal number of trials being classified to either state. This is likely not a valid assumption. Participants vary greatly in the proportion of time spent mind wandering[2,14], which this method fails to account for. These concerns motivate further search for a method of inferring attention fluctuations that captures variation in the degree and length of mind-wandering states.

Interestingly, analytic tools using Hidden Markov Models (HMM) provide ways to infer discrete internal states that avoid these pitfalls[50]. Specifically, Ashwood and colleagues[50] demonstrated that behavior from mice on a perceptual decision-making task could best be described in terms of transitions between distinct internal states that persist across many trials. The mice performed a simple task in which they had to respond to the location in which a stimulus appeared (left or right), but with varying evidence strength across trials. Discrete states were identified based on the relationship between the evidence strength of the stimulus and behavioral performance. Specifically, the modeling approach characterizes each state with a unique logistic Generalized Linear Model (GLM) that translates stimulus evidence into a response. Next, an HMM captures behavior in terms of transitions between these discrete, data-driven states. Ashwood and colleagues[50] found evidence that a GLM-HMM with one "engaged" state sensitive to stimulus strength and multiple "disengaged" states with response biases provided a superior fit of behavior relative to single-state psychophysical models, including those that model attentional lapses.

The GLM-HMM approach not only provides GLM parameters for each state, but also a set of transition probabilities between them. Combined with the behavior, these can be used to generate internal state predictions for every trial in the task. Ashwood and colleagues[50] validated these trial-level predictions by demonstrating that RT distributions and the rate of missed responses were qualitatively different between the engaged and disengaged states. Here, we describe a framework that combines this modeling approach with a task that adopts the best practices from the SART.

There have been previous approaches using HMMs to study human mind wandering. For example, Lee and colleagues[51] found that HMMs can be used to categorize participants based on eye movement patterns during the SART. Specifically, participants who showed eye movements with a stronger tendency to stay focused on the goal-relevant area of the computer screen showed increased behavioral performance and reduced self-reported mind wandering. Thus, eye movements can distinguish between groups of people employing different attentional strategies. Although this work shows the promise of using HMMs for detecting attentional states, this approach relies on specialized eye-tracking measures. In contrast, our approach infers changes in attentional states (and associated information processing strategies) directly from behavior, making it more scalable and easier to integrate across experimental contexts.

In another attempt, Bastian and Sackur[39] leveraged the difference in RT variability between attentional states, fitting a two-state HMM to participants' RT data on the SART. The main assumption behind this analysis is that because mind wandering leads to a different RT distribution than the on-task state, the model can estimate the posterior probability that the RT on each trial belongs to one of the two distributions. This allowed the authors to use the model to make predictions about the active state on each trial. Most importantly, they found that their model-predicted states indeed aligned with participants' subjective reports of mind wandering.

This approach demonstrates the potential of computational modeling in inferring attentional states. However, there are still notable shortcomings to this approach. First, this modeling framework does not validate the inferred states on data other than self-reported attentional states. This poses a problem when using this approach down the line in the intended way: without using thought probes. In that case, this framework leaves no way to validate the model-inferred states. Ideally, one should be able to validate inferred mind-wandering states from an HMM on some independent set of behavior (i.e., which it has not seen). For example, the temporal sequence of attentional state predictions could be used to examine a classic temporal effect where later trials are increasingly more likely to be categorized as "mind wandering" [52,53]. Second, this model assumes that the RT distributions of the two states have the same mean and only differ in their variance. This assumption is likely not valid, since an extensive set of studies indicates that RTs speed up when people are mind wandering[2,22,26,36]. A data-driven model with fewer between-state constraints should be better suited to capture behavioral differences between states. Finally, the duration of these inferred bouts of off- and on-task states was substantially shorter than those reported in the literature[54–56]. Together, these results leave an opportunity for the GLM-HMM framework to infer mind wandering from behavior alone while addressing these concerns.

One obstacle for current tasks (such as the SART) is that they do not generate behavioral patterns that are rich enough for validating states using an independent set of observations or for fitting GLMs. For example, the SART and the gradual continuous performance task[45] use categorical stimuli (either a go or no-go), which leads to the majority of trials being identical and likely producing similar behavior. Thus, if model-generated states were used to predict choice behavior, this would simply lead to a high probability of making the "go" response. Moving to a stimulus with continuously varying levels of evidence provides a solution for this issue. Importantly, it also allows for fitting a GLM-HMM on one behavioral measurement (choice) and validating the states in patterns of another (RT). Such validation would provide strong support for the ability of the model to tease apart meaningful cognitive states using HMMs. At the same time, even though the GLM-HMM framework has successfully captured attentional lapses in mice, this approach has not yet been used to study human mind wandering. One potential reason may be that the tasks used in this domain are not optimized to induce mind wandering in humans. Critically, mind-wandering tasks encourage habitual and repetitive responses. Thus, there is untapped potential for using the GLM-HMM approach to infer the temporal dynamics of mind wandering from behavior alone.

Here, we develop a paradigm (Fig. 1) that tackles these issues. The task (1) retains the repetitive response structure of mind-wandering tasks like the SART to induce mind wandering and (2) uses stimuli with continuously varying levels of evidence to allow the fitting of GLM-HMMs. In this perceptual decision-making task, participants indicate the direction of coherently moving dots in a random dot kinematogram (RDK) stimulus, with the proportion of coherence dots varying from trial to trial[57]. Importantly, 90% of trials involve the same ("dominant") motion direction, which encourages habitual and repetitive responding. We included occasional thought probes for validation. We hypothesized that this task, combined with the GLM-HMM, can be used to infer mind wandering without thought probes,

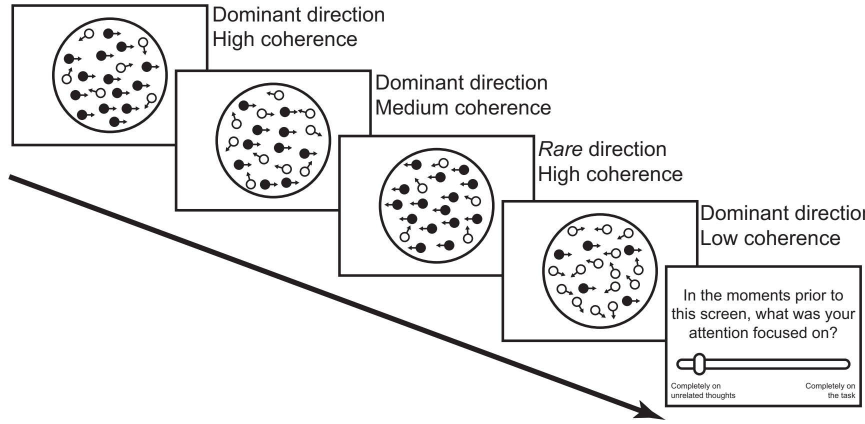

Fig. 1 | Task procedure. On every trial, participants see an RDK stimulus and indicate the coherent motion with key presses (P or Q). Participants also respond to thought probes throughout the task to report internal states. Even though coherence levels varied across trials, one "dominant" direction (here: right) was more likely (90%) to occur than the other "rare" direction (10%).

revealing a detailed insight into its temporal dynamics. In short, we find that this task successfully renders distinct behavioral profiles for two self-reported attentional states (on- vs. off-task), with participants relying more on the task's bias when they report being off-task. Next, we fit individual GLM-HMMs to task behavior, using these behavioral profiles to label the resulting states as on- and off-task. Critically, the inferred trial-wise state classifications from this modeling approach provided converging evidence that they tracked mind wandering: Model-derived off-task trials involved RT times and higher degrees of self-reported mind wandering. In addition, the modeling approach uncovered a classic temporal effect[22,52,53], with trials later in the task being more likely to be classified as off-task[45]. Together, these findings suggest that GLM-HMMs can be used to infer dynamics in attentional state switches from behavior alone.

Methods

This study was not preregistered.

Participants

We recruited 117 participants (33 males, 81 females, 3 non-binary; Age range = 18–22, M = 19.44, SD = 1.16) from the student population of Washington University in St. Louis. Information about sex was provided through participants' self-report. We did not collect data on race or ethnicity. Participants received academic course credit through participation. All participants completed the experiment in person and provided informed consent before starting. We excluded participants with accuracy rates lower than 2/3 from the analysis. We also excluded those with an average probe response greater than 90, since a high level of average focus may indicate that the participant did not engage in any mind-wandering behavior. To ensure a substantial level of attention fluctuation, we also excluded participants whose maximum and minimum probe responses had a difference of less than 30 units. This results in an exclusion of data from 24 participants, resulting in a sample size of $N = 93$. All procedures were approved by the Washington University in St. Louis Institutional Review Board.

Procedure

Participants in our task (Fig. 1) viewed and responded to a sequence of RDK stimuli[57–59]. Occasionally, a thought probe stopped the main task and asked participants to indicate their level of focus just before the probe.

Random dot motion task

Each trial started with a fixation cross in the center of the screen. Participants then saw an RDK stimulus, where a total of 200 dots moved within a round aperture with a width of 250 pixels. A proportion of the dots moved together towards a coherent direction, while the remaining dots moved in random directions. Participants were instructed to indicate whether the coherent dots were moving either left or right with one of two possible keyboard responses (Q and P).

Importantly, the task involved highly repetitive response requirements, modeled after the SART: 90% of the trials had the same coherent direction and therefore elicited the same correct response. This "dominant" direction was randomly chosen for each participant and stayed the same throughout the entire task.

To include evidence on a continuous scale, we varied the proportion of dots moving in the coherent direction. We refer to this proportion as the "coherence level". The coherence level was chosen randomly from 10 to 50% on each trial. To include more noise, the coherent direction was randomly chosen from the range of $[-30, 30]$ degrees, deviating from the left and right direction. We refer to this movement direction as "deviance". Both the coherence level and deviance were randomly chosen on each trial, independent of the correct response and dominant direction. The inter-trial interval was 1.5 s. The response deadline was 4 s. With 600 trials in total, the experiment lasted ~30 min.

Thought probes

To validate our experimental approach, we included occasional thought probes to obtain attentional states throughout the experiment. Before the experiment, participants received information about the terms "task-related" and "task-unrelated" thoughts meant, and how to respond to a thought probe. On average, a thought probe stopped the main task every 30 trials. To lower the interruption from probes on the main task, we chose this probe rate by selecting the lowest probe rate of 2%[38] in the literature we examined. Specifically, we matched the average probe interval by comparing average RTs between their task and our pilot study. Because we estimated focus levels on trials without a probe by performing linear interpolation on neighboring probe responses, we asked participants to report the focus of their attention on a continuous slider that ranges from "completely on unrelated thoughts" (0 units) to "completely on the task" (100 units). Participants were informed that their probe responses would not affect the credit they received and were encouraged to respond with complete honesty.

This experiment was developed with the jsPsych library[58]. We used the RDK plugin to display our stimulus[59].

Analysis

Interpolation of attentional state from self-report. To estimate the attentional states for trials without a thought probe from the self-report measures, we performed linear interpolation on every pair of two neighboring probe responses. To do this, we first took a pair of neighboring probe trials, where the trial is associated with an available probe response. We then assigned each of the unprobed trials between these two trials a calculated focus score by performing linear interpolation on the two responses. We assumed the first trial had a probe response of 100. We performed the same procedure for all pairs of neighboring probe trials. As a result, most trials in the experiment are associated with a focus score, except for the small number of trials after the last available probe response. For the Figures where measures are plotted as a function of self-

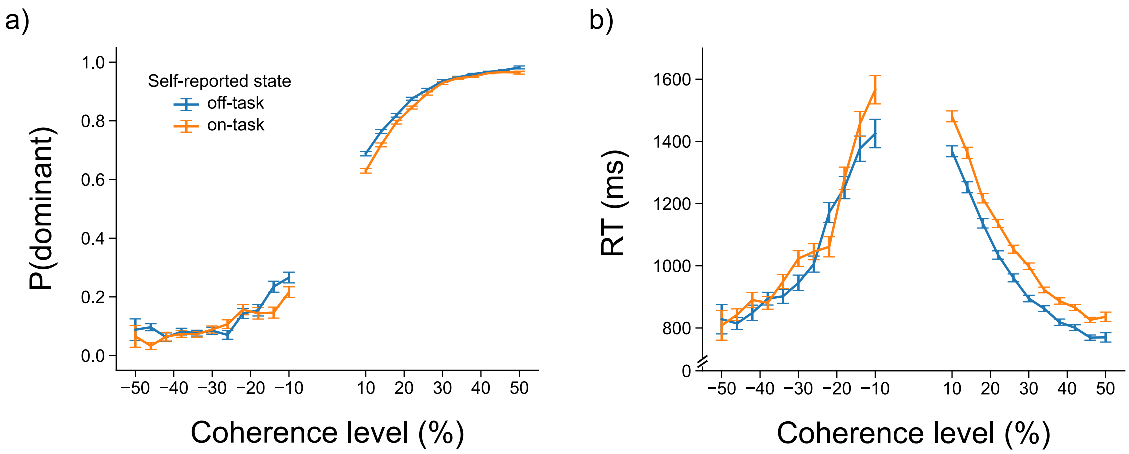

Fig. 2 | Behavioral measurements by coherence level and self-reported attentional state. a Probability of choosing the dominant response. **b** Response time. Error bars: standard error of the mean. Negative coherence levels on the x-axis represent trials with rare motion direction. $N = 93$ participants.

reported focus, we split all trials into two categories (on- vs. off-task) based on this variable. To do so, we normalized these focus scores with respect to each participant's mean score. The trials with normalized focus scores greater than zero were labeled "on-task" whereas the rest were labeled "off-task". Because probes were interspersed randomly throughout the trial sequence, two probes could potentially be quite far away from each other. Therefore, we only kept the trials that were at most within 20 trials from the closest preceding or following probes. However, such data exclusion is only relevant for analyzing observed behaviors with respect to self-reported states (Fig. 2). All remaining analyses incorporated all trials, regardless of their positions.

General linear mixed effects model. We first analyzed behavioral measures, including choice behavior and RT. To examine the behavioral patterns in different self-reported states, we fit a hierarchical logistic regression for choice behavior (P(dominant)). The response requirement in our design naturally guaranteed that choice data met the basic assumptions of logistic regression, where each observation is binary and independent of the others:

$$P(dominant) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * state + \beta_2 * coherence)}}$$

We included attentional state and coherence level as model predictors. In this regression, attentional state is the normalized self-reported focus score, and coherence is a continuous regressor, with values ranging from -50 to 50, and negative values indicating the rare motion direction. We also fit a hierarchical linear regression predicting RT on correct trials with the same predictors as above. RT distributions did not have normal distributions. Addressing this non-normality, we performed a log transformation for all analyses where RT was involved as a dependent variable:

$$log(RT) = \beta_0 + \beta_1 state + \beta_2 coherence$$

All model fitting was done with R package lme4[60]. Effect sizes and confidence intervals were calculated with R package effectsize[61] and r2glmm[62].

Hierarchical drift diffusion modeling. To examine how attentional states affect the perceptual decision making process, we fit a hierarchical drift diffusion model (HDDM)[63,64]. This model fits choice and RT data on coherence level and self-reported attentional state. All choices on priors followed the original work on the model package[64]. We estimated the drift rate (v) and initial bias (z). Drift rate was free to vary by coherence level and attentional state, and initial bias (z) was free to vary by attentional state. To estimate the HDDM parameters, 2000 samples from the posterior were generated with Monte Carlo methods, and 200 samples were discarded as burn-in. For within-subject effects, we calculated the difference in the fitted parameters for on-task trials and off-task trials and examined the posterior distributions of these differences. A full description of posterior distributions is available on OSF.

GLM-HMM

To infer participants' internal states based on the relationship between behavior and stimulus strength, we fit a HMM where each discrete state corresponds to a set of decision strategies[50]. For each participant, we fit GLM-HMM on the coherence level and choice behavior. The model fitting procedure aims to learn the transition matrix, the GLM weights influencing choice in each state, and the initial state distribution. We followed the same structure as the original work[50]. The prior over the GLM weights was an independent, zero-mean Gaussian. An independent Dirichlet prior was placed on each row of the transition matrix as well as the initial state distribution.

This set of parameters was collectively fit to choice data using maximum a posteriori estimation, practically implemented by the EM algorithm. The E-step uses a forward-backward algorithm. In each iteration in the forward pass, the model estimates the posterior state probabilities up until each trial t, given the current set of parameters. In the backward pass, the model calculates the posterior probability of the choice data beyond trial t. These probabilities are then used to calculate an estimate for the log-posterior of the parameters given the choice and input data. The M-step of the EM algorithm maximizes this estimate with respect to the model parameters. Initial state distributions and transition matrices were calculated through closed-form updates, whereas GLM weights are found numerically. As a result, the model estimated a set of weights placed on evidence strength and bias, respectively for each state, as well as transition probabilities between the states. The model also generated a prediction of the current state for each trial.

Model comparison. To identify the optimal model, we fit models with different numbers of latent states from one to four for each participant and perform model comparison. Note that a one-state GLM-HMM reduces to a basic logistic regression with only one intercept and one coefficient for each participant. To assess the performance of a model, we randomly withheld 10% of the trials as the testing set and fit the model on the remaining trials. After we obtained the fitted parameters, we provided the model with the complete dataset.

Then, based on the fitted parameters, the model estimated PP_{ts}, the posterior probabilities of being in state s, for each trial t. The model also estimated $Pdom_{ts}$, the probabilities of making the dominant response, assuming trial t is in state s. We multiplied these two probabilities for each state and summed over the resulting products across all states s to calculate the total probability of making the dominant response in each trial t:

$$Pdom\_total_t = \sum_s^S PP_{ts} Pdom_{ts}$$

where S is the set of all possible states. We then calculated predictive error by averaging the differences between actual responses and model predictions for each trial t:

$$error = \frac{1}{T} \sum_t^T |Pdom\_total_t - y_t|$$

where T is the set of all held-out trials, and y is the actual choice observed.

We performed such an analysis on 100 different randomly defined held-out sets, and we obtained the prediction error by calculating the average value of these errors. Model predictive accuracy is then defined as the complement of predictive error.

Validating model predictions of states. Based on the results from our model comparison analysis, we concluded that the two-state model is sufficient in capturing behavioral patterns and thus used the two-state model for the following analysis. According to results from both the hierarchical GLM and HDDM, the off-task state is associated with higher bias, irrelevant to stimulus strength. Thus, we labeled the state with a higher weight for bias to be "off-task", and the other state as "on-task".

To test the validity of model-generated predictions of attentional states, we performed several analyses. We first compared the self reports and model predictions over time. To do so, we calculated the average normalized focus scores for each trial throughout the task across participants. We performed the same analysis using model predicted probability of being on-task.

Next, we directly compared the model predictions and self reports on the individual subject level. For each participant, we calculated the average level of normalized self-reported focus scores for trials labeled "on-task" and "off-task" by the model, respectively. Then we calculated the difference in focus between the two types of trials for each participant.

Next, we examined the relationship between evidence strength and RTs using model-generated states. This analysis uses the same methods described above for examining behavioral patterns with different self-reported states. Lastly, we investigated whether model predictions correlate with two known markers of mind wandering, RT variability and speeding. To do this, we calculated the correlations between the proportion of trials categorized as "off-task" and (1) RT variability and (2) average RT across subjects. In contrast with the analyses reported in prior sections, these analyses were performed on all trials (without discarding those that were more than 20 trials away from any thought probe), which allows us to test model predictions for the entire task duration.

Effect of probes on attentional states. We examined whether self-report thought probes transiently altered the level of focused attention in our task. For each participant, we calculated the difference between the average probability of being on-task for trials after probes and for all other trials. To rule out that this effect only reflected simple time-on-task effects, we also ran a control analysis where changes in each participant's attentional states were calculated using the probe positions of other participants. That is, we computed a given participant's probability of being on-task between these types of trials as if the probes had occurred at the times experienced by another participant. This Monte Carlo permutation procedure preserved the overall temporal structure of the task while disrupting any direct relationship between a participant's own probes and their attentional state. We repeated this procedure 100 times for each participant (using randomly drawn probe positions from other participants on each iteration), generating a "null" distribution of 100 observations. We then computed the true effect against this distribution of permuted effects, which allowed us to assess whether the observed decline in attentional focus following probes exceeded what would be expected by chance.

Results

On each trial of our task (Fig. 1), participants ($N = 93$) responded to an array of moving dots (an RDK). A subset of "coherent" dots moved in the same direction (either left or right), and the remaining dots moved in random directions. The coherence of motion (the percentage of dots moving in the same direction) varied across trials. Participants were instructed to use a left/right key press to indicate the direction of the coherently moving dots.

To facilitate mind wandering, one randomly selected direction was more prevalent ("dominant", 90% of trials), and the other was less prevalent ("rare", 10%). This feature of the task introduces a repetitive response requirement like the SART, which allows participants to rely on the bias in coherent motion.

We included infrequent thought probes in the task to characterize the behavioral profile of attentional states and to validate our computational modeling framework. In these trials, participants reported the focus of their attention just before the probe on a continuous scale from "completely on the task" to "completely on unrelated thoughts". At face value, the continuous nature of this scale stands at odds with the assumption of our HMM framework that people shift between discrete states. However, the continuous nature of the self-report allows participants to indicate uncertainty about their cognitive state (using less extreme values), and they facilitate descriptive interpolation between probes to provide trial-by-trial comparison to the model-inferred states.

When off-task, participants make more and faster dominant responses

To analyze the behavioral profile of attentional states, we assigned each trial an attentional focus score by linearly interpolating between pairs of self-reported scores (not including trials that were removed more than 20 trials from the nearest probe). Note that this practice is only relevant for the initial set of validation results reported in this section (see "Methods" for details).

Next, we analyzed choice behavior using a hierarchical logistic regression model, predicting the probability of making the dominant response from the degree of evidence for dominant motion and the continuous focus score. Unsurprisingly, the probability of making the dominant response increased as the evidence for the dominant motion increased ($z = 94.79$, $p < 0.001$, OR = 1.09, 95% CI [1.08, 1.09]). Importantly, we saw a main effect of self-reported focus, with the probability of making the dominant response increasing as self-reported focus decreased ($z = -9.52$, $p < 0.001$, OR = 0.88, 95% CI [0.85, 0.90]). For visualization purposes, Fig. 2a shows these findings with self-reported state as a discretized categorical variable. All analyses reported here use the continuous version of this variable (though analyses using the categorical version offered similar results).

Figure 2b shows response times (RTs) on correct trials as a function of coherence and self-reported state. We saw an expected decrease in RT with increasing coherence for both dominant (positive values) and rare (negative values) motion. Because of the non-monotonic relationship between coherence level and RT, we fit GLMs for dominant and rare motion trials separately. These models predicted log-transformed RT (due to the non-normal nature of RT distributions) as a function of coherence and self-reported focus. For dominant trials, RTs decreased as absolute coherence increased ($t(42780) = -68.61$, $p < 0.001$, partial $R^2 = 0.098$, 95% CI [−0.014, −0.013]). RTs also increased when self-reported focus increased ($t(42798) = 25.07$, $p < 0.001$, partial $R^2 = 0.014$, 95% CI [−0.053, −0.062]). For rare trials, we again found coherence level ($t(4686) = -28.61$, $p < 0.001$, partial $R^2 = 0.189$, 95% CI [−0.016, −0.014]) and attentional state to be significant predictors ($t(4676) = 5.30$, $p < 0.001$, partial $R^2 = 0.008$, 95% CI [0.020, 0.043]). Finally, we directly compared the effect of focus on RT and found that it was larger for dominant compared to rare trials ($t(53999) = 2.29$, $p = 0.022$, partial 95% $R^2 < 0.001$, CI [0.006, 0.081]).

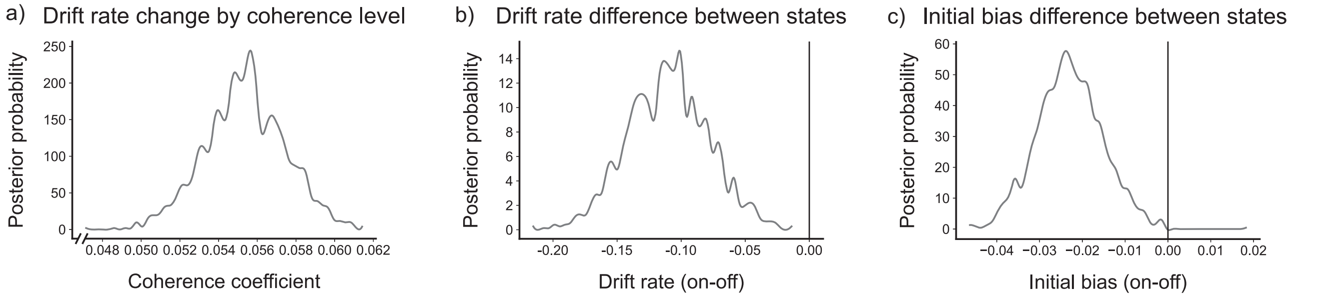

Fig. 3 | Distribution of the posterior probability of fitted parameters in relation to evidence strength and states. a Drift rate change by coherence level; **b** Drift rate difference between the two states; **c** Initial bias difference between the two states. Full description of posterior distributions is provided on OSF. $N = 93$ participants.

Together, these behavioral findings indicate that participants in the task draw more on the task's bias when off-task, regardless of the evidence, which leads them to make more correct and fast responses on dominant trials, but more errors on rare trials. In other words, they may start sampling stimulus evidence not from a neutral starting position, but from one in favor of the dominant direction.

Drift diffusion modeling of self-reported focus

This prediction is particularly well-suited to be tested using sequential sampling models such as the DDM. This class of computational models describes decision making as a process of sequential sampling of evidence for available responses until a threshold is reached. Thus, the DDM captures both RTs and choices.

For a formal test of our prediction, we fit a hierarchical DDM to analyze which of its parameters are affected by self-reported focus[63,64]. Model fitting was done using the HDDM package[64], which uses a Markov chain Monte Carlo sampling method to estimate posterior parameter distributions. This model predicted the combination of choice RT data as a function of coherence level and self-reported attentional state using Bayesian hierarchical regression. Specifically, we let the drift rate (v) parametrically vary as a function of both the coherence level and self-reported focus, and the starting point bias (z) as a function of self-reported focus.

As expected from decades of research in this domain, we found that an increase in coherence level was associated with increased drift rate. All samples of the posterior distribution of this effect (Fig. 3a) were greater than zero (95% CI [0.051, 0.059]).

More importantly, we also found that increased self-reported focus was associated with both a larger starting point bias (95% CI [−0.038, −0.008]) and drift rate (95% CI [−0.17, −0.05]) (Fig. 3b, c). As self-reported focus increased, both drift rate and starting point bias increased. For both these within-subject effects, all posterior distribution samples were negative. These results indicate that when participants report being off-task, they show a stronger initial bias towards the dominant response as well as increased evidence accumulation. This confluence of effects has been repeatedly reported in tasks with a bias in stimulus evidence[63,65,66].

Two-state GLM-HMM captures behavioral patterns

Using self-reported attentional focus, the analyses reported so far suggest that behavior on our task is generated by different internal states. However, the goal of our investigation was to demonstrate that we can infer these from behavior alone.

To do this, we fit multi-state GLM-HMMs[50] to participants' data. These models assume that behavior is generated by a set of distinct states, each corresponding to the psychometric choice curve. For each participant, this procedure estimates two GLM parameters representing the stimulus weight and bias parameter of the choice curve for each state, as well as the transition probabilities between states (for details, see "Methods"). These can be combined to obtain a prediction of internal states for all trials in the task.

To determine the optimal number of states, we computed predictive accuracy on held-out trials for GLM-HMMs with the number of latent states ranging from one to four (Fig. 4). The two-state model strongly outperformed the simple one-state logistic regression. On average, it raised predictive accuracy by 1.11%. As we increased the number of states from two to three and four, the model performance only improved slightly, with a 0.28% and 0.14% increase in accuracy. Thus, the two-state model sufficiently captures behavior while minimizing model complexity, and therefore we chose it for further analysis.

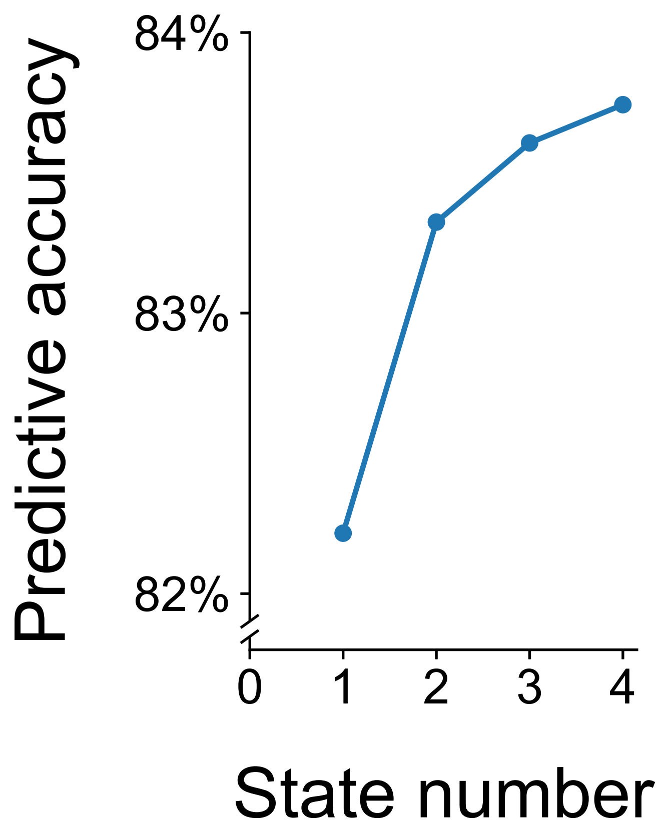

Fig. 4 | Model comparison results. Average predictive accuracy on the held-out trials for GLM-HMM models with different numbers of latent states.

Next, based on results from both the hierarchical GLM and HDDM, we labeled the two model-identified states using their associated state parameters. According to analyses, the off-task state was associated with an increased bias towards the dominant response. Thus, for each participant, we labeled the state with a higher bias parameter as "off-task", and the other as "on-task". Figure 5 shows the estimated transition matrix for one example participant (panel a), as well as the probability of the on-task state being active on each trial across the entire task (panel b). This participant was predicted to be on-task for 42% of the trials. Importantly, the state predictions suggest that this participant fluctuated between internal states throughout the entire task. Figure 5c shows a distribution of on-task state

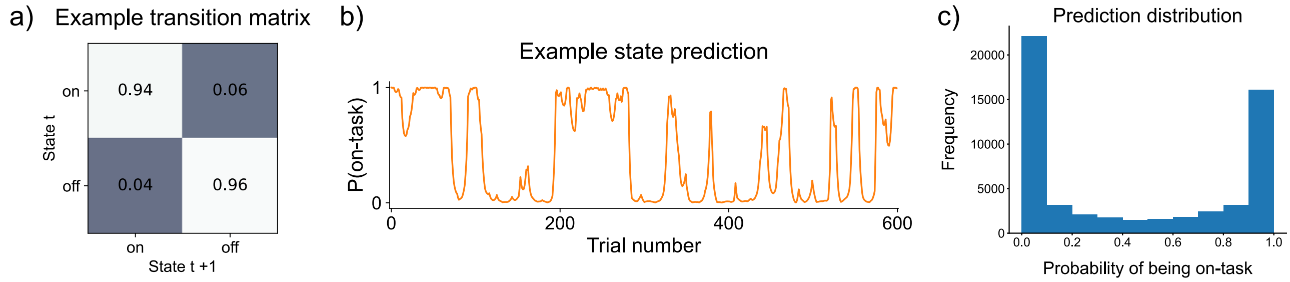

Fig. 5 | Model estimations for an example participant. a Transition probabilities between on-task and off-task states. **b** Model predicted probability of being on-task for all trials. **c** Distribution of model predicted states across all participants and trials.

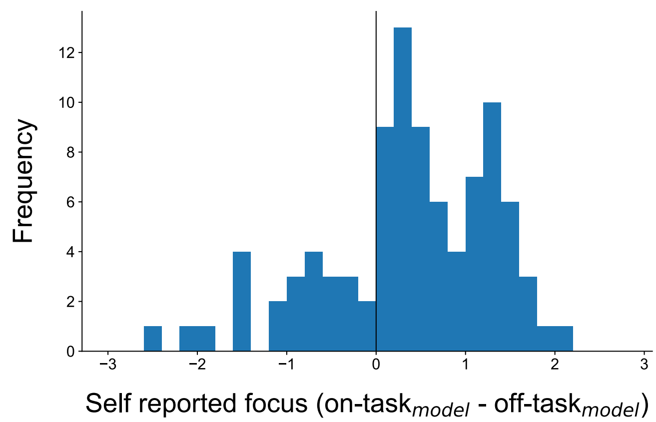

Fig. 6 | Distribution of difference in self-reported focus between states. This histogram shows the participant-level differences in self-reported focus (z-scored) between on-task and off-task trials based on model-predicted states. The vertical black line at zero indicates no difference.

probabilities for all participants across all trials, and the bimodality of this distribution indicates that the model has a high confidence in its state predictions.

Model-characterized attention fluctuations

We next examined the transition probabilities between the two states across participants. We found that the average probability of staying on-task was 0.96 (SD = 0.11) and the probability of staying off-task was 0.95 (SD = 0.08). Note that these two probabilities are very close to 1, indicating that both the on-task state and the mind-wandering state typically persist for several trials in a row. Even though they were close to each other, a paired Wilcoxon test showed that participants were significantly more likely to stay on-task than stay off-task ($z = -2.13$, $p = 0.03$, Cohen's $d = 0.45$, 95% CI [−0.014, −0.001]). This suggests that the on-task state is more stable than the off-task state.

While these results replicate the findings from Bastian and Sackur[39], the stay probabilities in our task are much higher than the ones reported in their work (0.89 and 0.82, respectively). In other words, our model predicts more stable states and fewer state switches. Accordingly, our model predicted longer durations for each on-task and off-task episode. On average, an on-task state persists for $1/(1 - 0.96) = 25.0$ trials, which translates to 26.4 s if accounting for the average RT of a trial. By the same logic, the expected duration of a sequence of off-task trials in the task was $1/(1 - 0.95) = 20.0$ trials or 21.1 s. These durations are substantially higher than those found using the RT-variability HMM procedure by Bastian and Sackur[39], which were 18.2 s and 11.1 s for on- and off-task, respectively. Indeed, previous research measuring the duration of mind-wandering epochs using a wide variety of methods[54–56] suggests that off-task episodes typically last 20 s or more. Our results align well with this set of convergent estimates, supporting the validity of our HMM-GLM framework in providing meaningful estimates of fluctuations in attentional engagement over realistic timescales.

Model-inferred attentional states track self-reported focus and behavioral performance

Finally, we set out to validate that the model-predicted states aligned with measures of behavior that were not used during model-fitting, such as participants' self-reported attentional focus and response times.

First, we used the fits to test whether participants indeed reported higher levels of focus on the trials that the model labeled as on-task. We found that participants' self-reported focus was higher on model-predicted on-task trials compared to off-task trials (Fig. 6, $t(92) = 3.52$, $p < 0.001$, Cohen's $d = 0.37$, 95% CI [0.15, 0.54]). We also examined the relation between self-reported and model-predicted focus on probe trials alone, without using the interpolated scores. A hierarchical linear regression showed that model predictions successfully predicted self-reported focus ($t(1731.23) = 9.90$, $p < 0.001$, partial $R^2 = 0.02$, 95% CI [6.37, 13.31]). These results indicate that the predicted attentional states from the GLM-HMM, which were fit using only stimulus strength and choice, aligned with self-reported mind wandering.

Second, we found that the average probability of the on-task state decreased throughout the task ($t(55705) = -74.0$, $p < 0.001$, partial $R^2 = 0.06$, 95% CI [−0.00060, −0.00057]; Fig. 7a). This finding is noteworthy because the GLM-HMMs do not formally capture such trends (e.g., by including a parameter that captures changes in transition probabilities over time). Thus, without assuming any underlying dynamics, the model reveals the widely observed trend of increased mind wandering over time. These results suggest that the analytic procedure was sensitive to a behavioral trend that influenced the inferred state probabilities. Importantly, an analogous effect was present for self-reported focus level ($t(54466) = -131.9$, $p < 0.001$, partial $R^2 = 0.24$, 95% CI [−0.0030, −0.0029]; Fig. 7b). Thus, both self-reports and model predictions indicate an overall decline in focus level as the task progressed, providing support for the validity of the model-predicted states. Such declines in attentional focus are a key characteristic of mind-wandering behavior[22,52,53]. Importantly, a follow-up mixed-effects model, regressing both model-predicted state, trial number, and their interaction (including random intercepts and slopes) on self-reported focus, indicated a main effect of trial number ($t(47.6) = -5.54$, $p < 0.001$, partial $R^2 = 0.12$, 95% CI [−0.003, −0.002]) and model-predicted state ($t(203.0) = 4.42$, $p < 0.001$, partial $R^2 = 0.02$, 95% CI [0.213, 0.872]), but not their interaction ($t(53968.0) = -0.09$, $p = 0.925$, partial $R^2 = 0.07$, 95% CI [−0.004, −0.001]). These results again show that self-reported focus decreases as trial number goes up, but, more importantly, the probability of being on-task (i.e., the model-predicted state) positively tracks self-reported focus beyond this general decline.

Thirdly, because our GLM-HMMs were not fit on RT data, we could examine patterns of RTs based on the model-predicted state categorization (Fig. 8). The trends in RT were very similar to those analyzed using self-reported focus states. Two mixed linear models fit on model-predicted states replicated the core results of the model fit using self-reported states. For both dominant and rare motion trials, both coherence level (dominant:

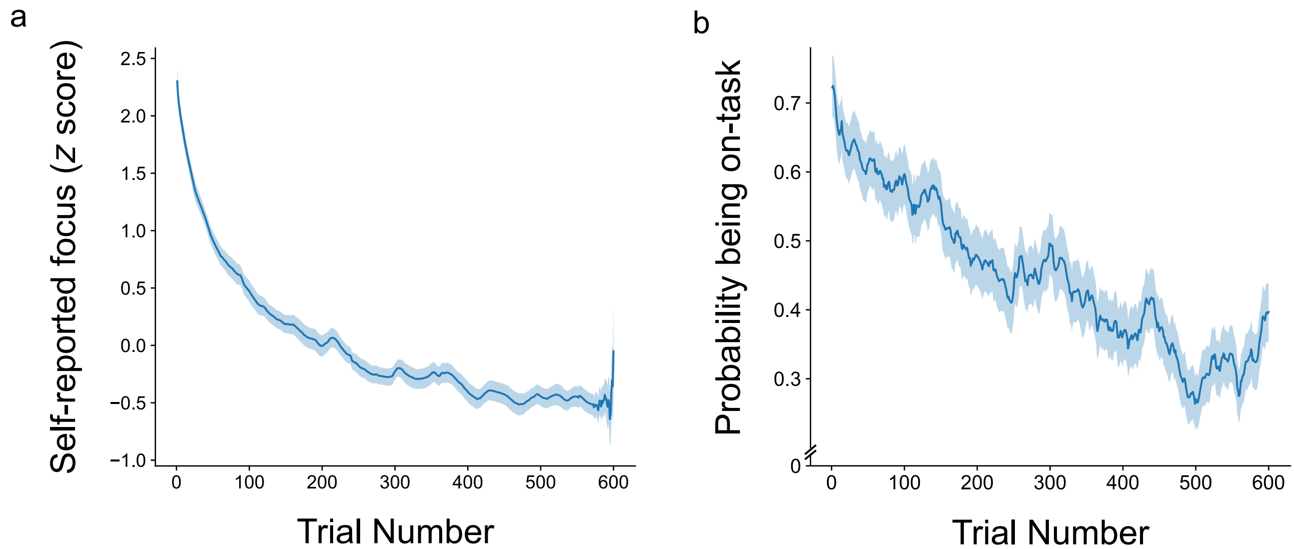

Fig. 7 | Time-on-task effects revealed by self-reported focus and model predictions. a Self-reported focus (*z* scores) over time, averaged across participants. **b** The model-predicted probability of being on-task over time, averaged across participants. Shaded area: standard error of the mean. $N = 93$ participants.

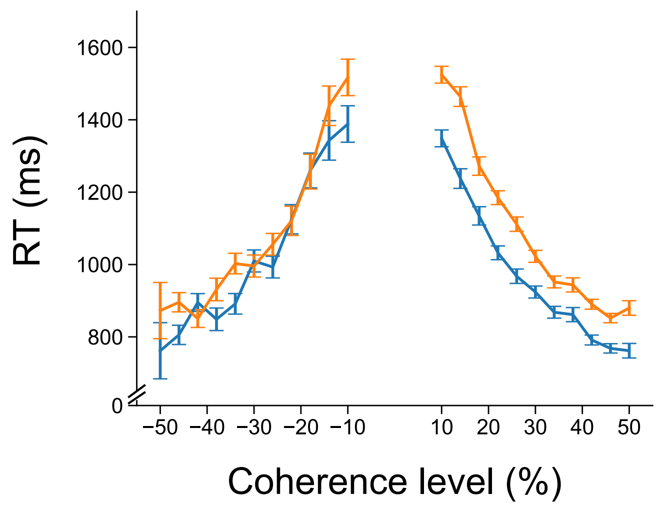

Fig. 8 | RT pattern using model predicted state categorization. Error bars: standard error of the mean. $N = 93$ participants.

$t(43872) = -70.11$, $p < 0.001$, partial $R^2 = 0.10$, 95% CI [−0.014, −0.013]; rare: $t(4806) = 28.04$, $p < 0.001$, partial $R^2 = 0.14$, 95% CI [16.77, 19.29]) and focus state (dominant: $t(43750) = 25.78$, $p < 0.001$, partial $R^2 = 0.01$, 95% CI [0.170, 0.198]; rare: $t(4722) = 2.12$, $p = 0.03$, partial $R^2 = 0.001$, 95% CI [3.70, 91.96]) were significant predictors for RT. Thus, this analysis provides further corroboration that the model-predicted states align with self-reported focus levels (Fig. 2b). In addition, we also fit a hierarchical linear model that predicted RTs across all trials as a function of trial type (dominant vs rare), absolute value of coherence level, and the model-predicted probability of being on task. In addition to main effects of coherence ($t(48753) = -47.10$, $p < 0.001$, partial $R^2 = 0.02$, 95% CI [−0.014, −0.012]), model-predicted on-task probability ($t(48816) = 14.77$, $p < 0.001$, partial $R^2 = 0.001$, 95% CI [0.208, 0.271]) and trial type ($t(48753) = 3.72$, $p < 0.001$, partial $R^2 < 0.001$, 95% CI [0.054, 0.174]), this revealed a significant interaction between trial type and model-predicted focus level ($t(48752) = -2.46$, $p = 0.01$, partial $R^2 < 0.001$, 95% CI [−0.208, −0.024]). In other words, reduced attentional focus led to speeded responses for the dominant trials, but this effect was less distinct for correctly answered rare trials where participants successfully overcame their response bias. These RT results provide convergent external validation that the GLM-HMMs inferred fluctuations in attentional state.

Characterizing the transition dynamics of group-level differences in mind wandering

After validating the model at the within-subject level, we aimed to investigate between-subject effects. First, we leveraged the across-participant variation in the proportion of on-task trials to determine the mechanistic nature of increased mind wandering. To do this, we adopted an extreme-group analysis, comparing the estimated state transition probabilities between the participants with the 25% highest and lowest proportion of on-task trials (Fig. 9). This analysis revealed that the group with the highest rates of mind wandering showed a reduced probability of transitioning from the on- to the off-task state (i.e., reduced lapsing), $t(23) = -2.74$, $p = 0.012$, $d = -0.57$, 95% CI [−0.066, −0.009], and an increased probability of transitioning from the off- to the on-task state (reengagement), $t(23) = 3.40$, $p = 0.002$, $d = 0.71$, 95% CI [0.049, 0.203]. This shows that participants in the on-task group were less likely to lapse and more likely to get back on task after lapsing.

Individual differences in mind wandering relate to response time variability and speeding

Next, we examined whether individual differences in model-predicted mind wandering aligned with known behavioral markers of attentional lapses, such as response time variability[45–49] and speeding[22,36].

First, we investigated whether participants' RT variability correlated with model-predicted rates of mind wandering. Interestingly, we found a negative correlation between RT variability and the proportion of trials categorized as "off-task" on the subject level (Fig. 10a) ($r(91) = -0.28$, $p = 0.006$, 95% CI [−0.46, −0.08]). This result seems to contradict prior work, using the SART or related tasks, in which mind wandering is associated with increased RT variability[45,46]. In our task, however, we believe that the inverse pattern is to be expected because there are trial-to-trial fluctuations in the degree of perceptual evidence (Fig. 2): if participants are off-task, the increased response bias should make participants less variable, especially on the more numerous dominant trials.

Moreover, we also found a negative correlation between average RT and proportion of off-task trials (Fig. 9b) ($r(91) = -0.21$, $p = 0.04$, 95% CI [−0.40, −0.01]). This result, which indicates increased speeding for participants with a higher prevalence of off-task trials, is more directly in line with prior research[22,36,67].

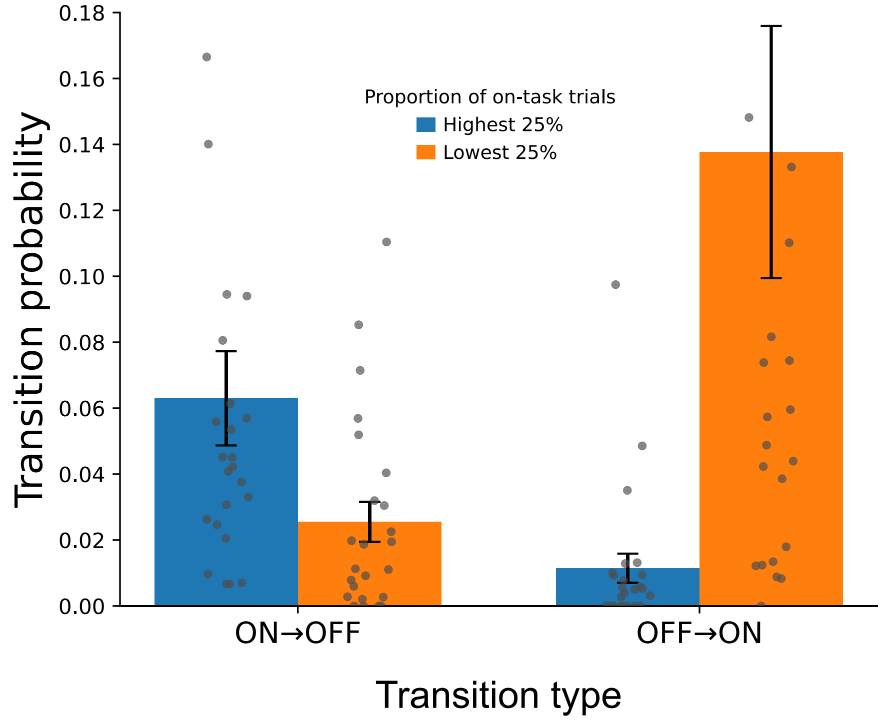

Fig. 9 | State transition probabilities vary between the groups with the 25% highest (orange) and lowest (blue) proportions of on-task trials. Reduced mind wandering is driven by both increased reengagement and reduced lapsing. Error bars: standard error of the mean. $N = 24$ participants in each group.

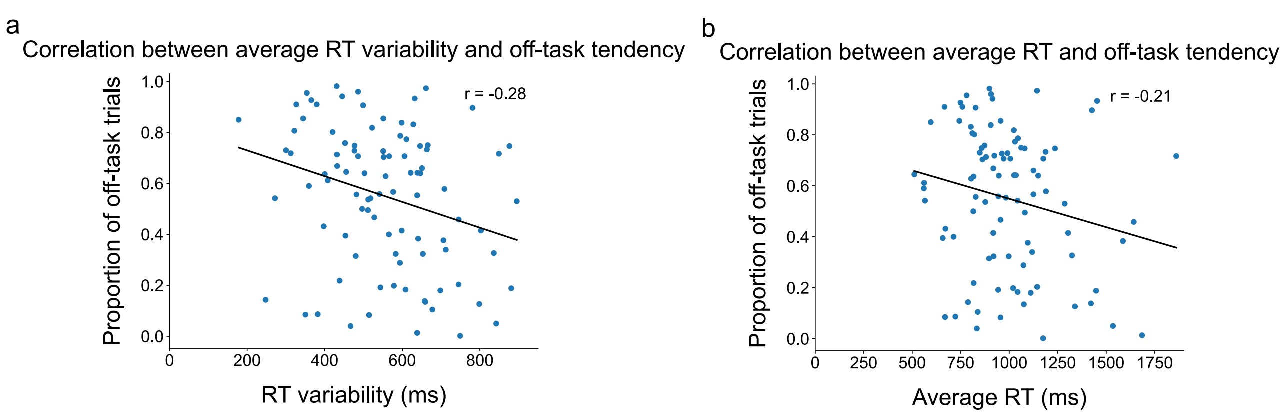

Fig. 10 | Subject-level correlation between known RT correlates and off-task tendency, measured by proportion of off-task trials. a Average RT variability. **b** Average RT. $N = 93$ participants.

As a next step, we investigated whether these correlations were primarily driven by participants' propensity to reengage (probability of transition of switching from the off-task to the on-task state) or to stay on-task (on-task to on-task). We found a significant correlation between the probability to stay on-task and RT variability, $r(91) = 0.25$, $p = 0.018$, 95% CI [−0.43, −0.04] and speeding, $r(91) = 0.21$, $p = 0.048$, 95% CI [−0.39, −0.00]. In contrast, we saw no such correlations for the transition probability from off- to on-task for neither RT variability ($r(91) = 0.11$, $p = 0.29$, 95% CI [−0.10, 0.31]) nor speeding ($r(91) = 0.10$, $p = 0.36$, 95% CI [−0.11, 0.29]).

Intrusive thought probes transiently decrease attentional focus

One of the motivations for developing a method to detect mind wandering without thought probes is to minimize their potential interruption on the natural attentional fluctuations. A key advantage of our modeling approach is that it estimates the probability of being on-task for each trial, independent of thought probes, allowing us to directly examine whether probes themselves alter attentional states.

To test this, we compared the model-predicted probability of being on-task for trials immediately following a probe against all other trials. The predicted probability of being on-task was significantly lower on trials directly after a probe ($t(92) = -2.52$, $p = 0.03$, Cohen's $d = -0.26$, 95% CI [−0.021, −0.003]), suggesting that probes temporarily decrease attentional focus. However, because trials following a probe occur slightly later than all other trials, one alternative explanation is that this effect is driven by the general decrease in on-task trials over time (Fig. 6). We ran a control analysis to rule out this explanation. As detailed in the Methods, we used a Monte Carlo permutation test to compute each participant's effect using probe positions of all other participants, reasoning that those would capture the general decline in on-task trials over time but not probe-specific effects. As can be seen in Fig. 11, this analysis indicated no reliable effect in the bootstrapped data using yoked probe positions ($t(99) = -1.55$, $p = 0.25$, Cohen's $d = -0.12$, 95% CI [−0.0014, 0.0004]), and a significant effect ($p = 0.02$, percentile = 1%, two-tailed) when using participants' own positions.

The pattern may reflect a specific form of intentional mind wandering, in which participants strategically choose when to disengage from the task to

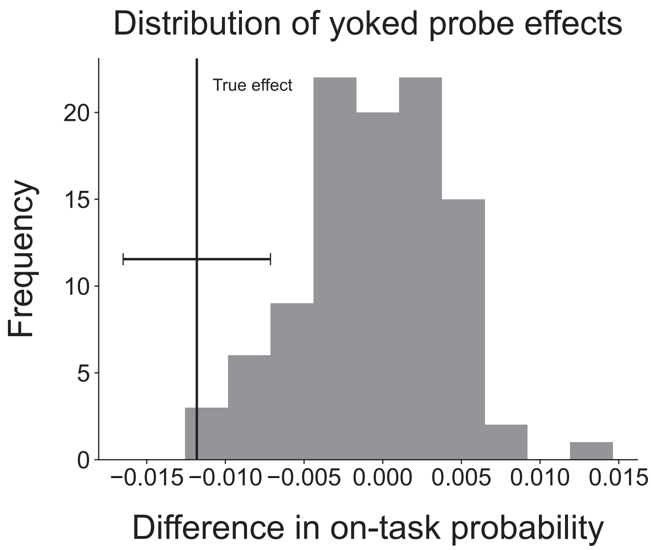

Fig. 11 | Histogram of random effects distribution generated from yoked probe positions (gray) compared with the observed mean probe effect (black line). Error bar: standard error of the mean. $N = 93$ participants.

process internal thoughts[37]. Because probes were rare, participants may have anticipated a period of uninterrupted task performance following a probe, allowing them "safely" redirect attention away from the task and rely more heavily on the response bias[68].

In summary, the results in this section indicate that GLM-HMMs, combined with a biased dot motion task, provide an opportunity to measure fluctuations in mind wandering from behavior alone.

Discussion

People need to sustain focused attention to achieve goals[26,69]. Yet, attention often lapses, as minds wander toward TUTs[5]. To study these dynamics in mental states, researchers typically use intermittent thought probes, which explicitly ask if thoughts are task-related. However, probes are rare and interrupt behavior[40]. Other researchers use response time variability as a proxy for attentional state, but this method does not account for between-person variability in mind wandering. Therefore, we developed an adaptation of a random dot motion task that, combined with an HMM, allows for the inference of within- and between-participant fluctuations in mind wandering from performance alone. As in regular dot motion tasks, participants indicate the coherent direction of each stimulus. However, the task adopts the repeated response structure from the traditional SART: 90% of trials exhibit the same motion direction, which encourages a habitual dominant response.

Because our task included conventional thought probes, we were able to demonstrate that it encourages a disengaged mind-wandering state, where participants rely more on the bias of the dominant response and less on trial-specific stimulus processing. This conclusion is further supported by the increase in mind wandering across the duration of the task[22,52,53], and increased commission errors on rare trials when participants reported being off-task[26,32,70,71].

Most importantly, this task allowed us to infer shifts between attentional states without relying on self-reported focus. We fit behavior on this task to a two-state GLM HMM, which provides inferred state probabilities (one labeled as off-task, the other as on-task) for each trial in the task. We provide three validations that these model-predicted states track shifts between attentional states. First, the latent states generated from this model predicted self-reported focus levels. Second, these state predictions successfully captured an increase in mind wandering over the course of the task, even though the GLM-HMM does not naturally capture such trends. Third, even though the modeling approach was only fit to choice behavior, the inferred mind-wandering trials showed faster RTs on common trials (a characteristic of reduced attention[22,26]). These results demonstrate that our approach measures attentional states and their temporal dynamics without relying on thought probes, paving the way for future mind-wandering research.

Our framework carries several advantages over conventional mind-wandering methodology. Most obviously, by allowing the measurement of mind wandering without thought probes, the task avoids any concerns around the confounding effect of probes themselves. This eliminates issues about how specific probe content and framing affect self reports[34,44]. Moreover, this also provides us with state inferences for all trials in the task. Unlike methods relying on RT variability as a proxy for lapses[45,46], our approach characterizes both within-person shifts in attentional states over time and between-person variation in the proportion of time spent in a disengaged state. In addition, while prior work also used an HMM approach to infer mind wandering[39], the states from that model could only be validated using subjective self-reports. By combining a GLM-HMM with a parametric stimulus, we were able to validate the attentional states using extraneous behavioral measurements, both at the trial level (RT speeding under wandering) and across the entire task (shift towards off-task state over time). We believe that we were able to uncover these patterns because our approach does not constrain how parameters vary between states (unlike the prior HMM approach), making the model fitting fully data-driven and maximally flexible. Notably, our modeling framework returned expected durations of the off- and on-task states that were more comparable with those reported in studies using physiological measurements[54,55]. Together, these findings indicate that the GLM-HMM approach carries several benefits over the prior method.

These features of our methodology can invigorate research on the consequences of mind wandering over time. By capturing fluctuations in attention from trial to trial, researchers can avoid relying on coarse groupings of trials based on certain events such as "trials before an error" or "trials preceding a probe"[18,72–74]. Other groups have used variability in RTs to predict attentional states[45,46,48]. While this provides state classifications for all trials, it requires researchers to set an arbitrary cut-off point for labeling them on- or off-task. In contrast, our approach naturally allows for the degree of mind wandering to vary across people, while generating predictions for all trials with no unused data.

We would like to point out that the GLM-HMMs were fit only on choice behavior and did not incorporate RT. However, RTs are a crucial source of variance in perceptual decision making tasks[75]. In fact, there is a rich literature that uses evidence sampling models (EAM) to jointly fit choice and RT[63,76]. This introduces the opportunity to develop an EAM-HMM approach, where each state is characterized by parameters for the sampling model instead of the GLM. In fact, a recent study by Kucharsky and colleagues[77] has taken promising steps in this direction, incorporating a simplified linear ballistic accumulator within an HMM. This approach was able to uncover dynamics in how participants were instructed to perform a simple cognitive task (focusing on accuracy vs. speed), demonstrating the feasibility of combining response time and choice data in an HMM and using this to infer latent cognitive parameters. Future work could extend these ideas to more fully parameterized sequential sampling models, such as a DDM. Compared to our current model, a DDM-HMM could increase sensitivity in state inference, as it would be based on both critical measurements of the decision-making process.

We believe that several aspects of our modeling approach can provide valuable perspectives on existing questions in the field of attention research. First, our task offers higher flexibility in identifying discrete attentional states engaged during task performance. The prevalent model of mind wandering is the lapse model, which assumes a weakened relationship between behavior and stimulus strength when attention lapses[78,79]. However, some researchers have suggested there are multiple distinct ways in which the mind can wander[37,80]. For example, some studies distinguish between off-task states based on whether awareness is involved, using more specific options such as "tuned out" (mind wandering with awareness) and "zoned out" (mind wandering without awareness) when acquiring self-reported states[5,18,81]. In fact, people show different patterns of behavior and brain activity when mind wandering with and without awareness, indicating these reflect distinct attentional states[82,83]. So far, such differences are not reflected in more precise computational models. The classic lapse model, which assumes two discrete states, likely falls short in accounting for this complexity. Our framework, on the other hand, can fit a custom number of discrete states and examine the differences in the specific components of the cognitive processes involved.

Our framework is also well-suited to studying mindfulness, the ability to maintain meta-cognitive awareness of the contents of one's own mind[84,85]. Extensive evidence suggests that increases in mindfulness are associated with better performance on tasks that require sustained attention[84,86–88]. This increase in performance may be caused by an increased ability to detect mind wandering and a subsequent diversion to on-task thoughts[88,89]. Alternatively, it may be related to an increased ability to maintain focus on the present experience[85,86]. Because our computational framework provides transition probabilities between attentional states, it can distinguish between these two hypotheses. If mindfulness improves sustained attention, one would expect a lower probability of switching from on-task states to off-task states in experienced practitioners. If mindfulness improves people's ability to bring their minds back to task-related thoughts when realizing they were mind wandering, one would expect a higher probability of switching from off-task states back to on-task states. Our modeling approach enables direct comparisons for properties related to attentional dynamics.

Finally, our computational approach holds promise to both inform neuroscientific theories of mind wandering, as well as to be improved by incorporating neural signals. An obvious target would be the default mode network (DMN), a network of brain regions spanning frontal and parietal cortex[90] that has been shown to play an important role in mind wandering[91–93]. For example, functional magnetic resonance imaging research has shown that lapses of attention are linked to increases in DMN activation[94,95], and that a reduced level of neural activation in DMN is associated with focused states compared to mind wandering states[96]. However, this body of work relies either on self-reported attentional focus or RT variability to index attentional states. The HMM framework introduced here can be used to estimate the neural correlates of attentional states on a trial-to-trial basis without thought probes, and while allowing the proportion of disengaged trials to vary between participants. Investigating the difference in neural responses between trials with different model-predicted states would provide a converging, and statistically more powerful, way to investigate the role of DMN in reduced attentional control. Moreover, trial-by-trial changes in DMN activity can then be incorporated into the HMM models to provide further demarcation of cognitive states.

A similar set of analyses can be done using electroencephalography (EEG), which has also been used to examine neural signals related to attentional states. A large body of work suggests that fluctuations in spectral power of alpha (Hz) and beta (Hz) frequency bands track temporal dynamics of attentional states[72,74,97–99]. Thus, it is worth characterizing how these markers relate to the model-predicted states from the HMM approach. Given the temporally precise prediction of our computational approach, we can also investigate neural responses on trials that involve state switches (either from or towards an on-task state). The direction of the motion in the stimuli in our task can be decoded from EEG signals[100,101]. This can also be leveraged to characterize the model-inferred states in terms of their neural implications. Specifically, because participants rely more on the task's bias on off-task trials, we predict reduced performance of motion direction decoders on these trials.

Limitations

We recognize that our framework may not be the optimal choice for some questions regarding the nature of mind wandering. Because we aim to infer mind wandering without using thought probes, our framework naturally fails to offer much information on the intentionality of mind wandering or the specific content of TUTs. For example, some researchers have highlighted the distinction between mind wandering with and without intentionality[38,102]. Studies on the content of TUTs offer probe response options that distinguish the specific topic, such as asking whether mind-wandering content is about the future or the past[103,104]. Distinctions are also made between inattention caused by self-initiated mind wandering or by external distraction[105,106]. These distinctions are crucial in forming a theoretical framework for mind wandering, but it is not obvious how our paradigm could capture these differences from behavioral measurements.

In addition, while we aimed to minimize the researcher degrees of freedom, we still had to choose how to interpolate focus levels on trials without a probe and which trials to discard during behavioral analysis. We recognize that these are also considered researcher degrees of freedom. Fortunately, these decisions were only necessary for validating our approach. Moreover, we emphasize that linear interpolation of probe reports was not intended as a model of attentional dynamics, which we assume to be discrete. All inferences about state transitions are derived exclusively from the GLM-HMM fit to behavioral data. Our results suggest that researchers can start using our framework to infer attentional states solely from behavioral measurements and without having to make similar design choices.

In conclusion, we developed a perceptual decision-making task to infer mind wandering purely from perceptual decision making. We found that participants' choices were more biased by the task structure when they reported being more off-task. This behavioral difference between attentional states allowed us to use computational modeling to capture the temporal dynamics of attention without the restriction and interruption from thought probes. Along with the great flexibility in stimulus space, our paradigm provides great potential in investigating how people's attention fluctuates in response to tasks, goals, and environments. We hope this paradigm will contribute to the psychological theories of mind wandering and attention.

Data availability

We provide access to all data collected on OSF.

Code availability

We provide access to all analysis scripts on OSF.

Received: 6 February 2025; Accepted: 6 February 2026;
Published online: 20 February 2026

References

1. Killingsworth, M. A. & Gilbert, D. T. A wandering mind is an unhappy mind. *Science* **330**, 932 (2010).
2. McVay, J. C. & Kane, M. J. Conducting the train of thought: working memory capacity, goal neglect, and mind wandering in an executive-control task. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 196–204 (2009).
3. Kane, M. J. & McVay, J. C. What mind wandering reveals about executive-control abilities and failures. *Curr. Dir. Psychol. Sci.* **21**, 348–354 (2012).
4. Giambra, L. M. A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Conscious. Cogn.* **4**, 1–21 (1995).
5. Smallwood, J., Beach, E., Schooler, J. W. & Handy, T. C. Going AWOL in the brain: mind wandering reduces cortical analysis of external events. *J. Cogn. Neurosci.* **20**, 458–469 (2008).
6. Smallwood, J. & Schooler, J. W. The restless mind. *Psychol. Bull.* **132**, 946–958 (2006).
7. Hawkins, G. E., Mittner, M., Forstmann, B. U. & Heathcote, A. Self-reported mind wandering reflects executive control and selective attention. *Psychon. Bull. Rev.* **29**, 2167–2180 (2022).
8. Yanko, M. R. & Spalek, T. M. Driving with the wandering mind: the effect that mind-wandering has on driving performance. *Hum. Factors* **56**, 260–269 (2014).
9. Baldwin, C. L. et al. Detecting and quantifying mind wandering during simulated driving. *Front. Hum. Neurosci.* **11**, 406 (2017).

10. Burdett, B. R. D., Charlton, S. G. & Starkey, N. J. Mind wandering during everyday driving: an on-road study. *Accid. Anal. Prev.* **122**, 76–84 (2019).
11. 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 (2018).
12. Kopp, K., D'Mello, S. & Mills, C. Influencing the occurrence of mind wandering while reading. *Conscious. Cogn.* **34**, 52–62 (2015).
13. Lindquist, S. I. & McLean, J. P. Daydreaming and its correlates in an educational environment. *Learn. Individ. Differ.* **21**, 158–167 (2011).
14. Deng, Y.-Q., Li, S. & Tang, Y.-Y. The relationship between wandering mind, depression and mindfulness. *Mindfulness* **5**, 124–128 (2014).
15. Seli, P., Beaty, R. E., Marty-Dugas, J. & Smilek, D. Depression, anxiety, and stress and the distinction between intentional and unintentional mind wandering. *Psychol. Conscious. Theory, Res., Pract.* **6**, 163 (2019).
16. Franklin, M. S. et al. Tracking distraction: the relationship between mind-wandering, meta-awareness, and ADHD symptomatology. *J. Atten. Disord.* **21**, 475–486 (2017).
17. Seli, P., Smallwood, J., Cheyne, J. A. & Smilek, D. On the relation of mind wandering and ADHD symptomatology. *Psychon. Bull. Rev.* **22**, 629–636 (2015).
18. Seli, P., Cheyne, J. A. & Smilek, D. Wandering minds and wavering rhythms: linking mind wandering and behavioral variability. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1–5 (2013).
19. Esterman, M., Rosenberg, M. D. & Noonan, S. K. Intrinsic fluctuations in sustained attention and distractor processing. *J. Neurosci.* **34**, 1724–1730 (2014).
20. Rosenberg, M. D., Finn, E. S., Constable, R. T. & Chun, M. M. Predicting moment-to-moment attentional state. *Neuroimage* **114**, 249–256 (2015).
21. Braboszcz, C. & Delorme, A. Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage* **54**, 3040–3047 (2011).
22. 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 (1997).
23. Smallwood, J., McSpadden, M. & Schooler, J. W. The lights are on but no one's home: meta-awareness and the decoupling of attention when the mind wanders. *Psychon. Bull. Rev.* **14**, 527–533 (2007).
24. McVay, J. C., Kane, M. J. & Kwapil, T. R. Tracking the train of thought from the laboratory into everyday life: an experience-sampling study of mind wandering across controlled and ecological contexts. *Psychon. Bull. Rev.* **16**, 857–863 (2009).
25. Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A. & Smilek, D. Anatomy of an error: a bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition* **111**, 98–113 (2009).
26. Manly, T., Robertson, I. H., Galloway, M. & Hawkins, K. The absent mind: further investigations of sustained attention to response. *Neuropsychologia* **37**, 661–670 (1999).
27. Robison, M. K., Miller, A. L. & Unsworth, N. Examining the effects of probe frequency, response options, and framing within the thought-probe method. *Behav. Res. Methods* **51**, 398–408 (2019).
28. Csikszentmihalyi, M. & Larson, R. Validity and reliability of the experience-sampling method. *J. Nerv. Ment. Dis.* **175**, 526–536 (1987).
29. Smallwood, J. & Schooler, J. W. The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).
30. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. & Stone, A. A. A survey method for characterizing daily life experience: the day reconstruction method. *Science* **306**, 1776–1780 (2004).
31. Weinstein, Y. Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behav. Res. Methods* **50**, 642–661 (2018).
32. Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M. & D'Argembeau, A. Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. *Acta Psychol.* **136**, 370–381 (2011).
33. Wiemers, E. A. & Redick, T. S. The influence of thought probes on performance: does the mind wander more if you ask it?. *Psychon. Bull. Rev.* **26**, 367–373 (2019).
34. Weinstein, Y., De Lima, H. J. & van der Zee, T. Are you mind-wandering, or is your mind on task? The effect of probe framing on mind-wandering reports. *Psychon. Bull. Rev.* **25**, 754–760 (2018).
35. Stawarczyk, D., Majerus, S., Catale, C. & D'Argembeau, A. Relationships between mind-wandering and attentional control abilities in young adults and adolescents. *Acta Psychol.* **148**, 25–36 (2014).
36. Smallwood, J. et al. Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. *Conscious. Cogn.* **13**, 657–690 (2004).
37. Seli, P., Risko, E. F., Smilek, D. & Schacter, D. L. Mind-wandering with and without intention. *Trends Cogn. Sci.* **20**, 605–617 (2016).
38. Seli, P., Risko, E. F. & Smilek, D. On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychol. Sci.* **27**, 685–691 (2016).
39. Bastian, M. & Sackur, J. Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. *Front. Psychol.* **4**, 573 (2013).
40. Schubert, A.-L., Frischkorn, G. T. & Rummel, J. The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychol. Res.* **84**, 1846–1856 (2020).
41. Vinski, M. T. & Watter, S. Priming honesty reduces subjective bias in self-report measures of mind wandering. *Conscious. Cogn.* **21**, 451–455 (2012).
42. Schooler, J. W. et al. Meta-awareness, perceptual decoupling and the wandering mind. *Trends Cogn. Sci.* **15**, 319–326 (2011).
43. Chu, M. T., Marks, E., Smith, C. L. & Chadwick, P. Self-caught methodologies for measuring mind wandering with meta-awareness: a systematic review. *Conscious. Cogn.* **108**, 103463 (2023).
44. Seli, P. et al. How pervasive is mind wandering, really?. *Conscious. Cogn.* **66**, 74–78 (2018).
45. Rosenberg, M. D., Noonan, S., DeGutis, J. & Esterman, M. Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Atten. Percept. Psychophys.* **75**, 426–439 (2013).
46. Esterman, M., Noonan, S. K., Rosenberg, M. D. & Degutis, J. In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* **23**, 2712–2723 (2013).
47. Weissman, D. H., Roberts, K. C., Visscher, K. M. & Woldorff, M. G. The neural bases of momentary lapses in attention. *Nat. Neurosci.* **9**, 971–978 (2006).
48. deBettencourt, M. T., Norman, K. A. & Turk-Browne, N. B. Forgetting from lapses of sustained attention. *Psychon. Bull. Rev.* **25**, 605–611 (2018).
49. deBettencourt, Keene, P. A., Awh, E. & Vogel, E. Real-time triggering reveals concurrent lapses of attention and working memory. *Nat. Hum. Behav.* **3**, 808–816 (2019).
50. Ashwood, Z. C. et al. Mice alternate between discrete strategies during perceptual decision-making. *Nat. Neurosci.* **25**, 201–212 (2022).
51. Lee, H.-H., Chen, Z.-L., Yeh, S.-L., Hsiao, J. H. & Wu, A.-Y. A. When eyes wander around: mind-wandering as revealed by eye movement analysis with hidden Markov models. *Sensors* **21**, 7569 (2021).

52. Zanesco, A. P., Denkova, E. & Jha, A. P. Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review. *Psychol. Bull*. https://doi.org/10.1037/bul0000424 (2024).
53. McVay, J. C. & Kane, M. J. Drifting from slow to "D'oh!": working memory capacity and mind wandering predict extreme reaction times and executive control errors. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 525–549 (2012).
54. Zheng, Y., Wang, D., Zhang, Y. & Xu, W. Detecting mind wandering: an objective method via simultaneous control of respiration and fingertip pressure. *Front. Psychol.* **10**, 216 (2019).
55. Grandchamp, R., Braboszcz, C. & Delorme, A. Oculometric variations during mind wandering. *Front. Psychol.* **5**, 31 (2014).
56. Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E. & Barsalou, L. W. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *Neuroimage* **59**, 750–760 (2012).
57. Williams, D. W. & Sekuler, R. Coherent global motion percepts from stochastic local motions. *Vis. Res* **24**, 55–62 (1984).
58. de Leeuw, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
59. Rajananda, S., Lau, H. & Odegaard, B. A random-dot kinematogram for web-based vision research. *bioRxiv* 192377. https://doi.org/10.1101/192377 (2017).
60. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
61. Ben-Shachar, M., Lüdecke, D. & Makowski, D. Effectsize: estimation of effect size indices and standardized parameters. *J. Open Source Softw.* **5**, 2815 (2020).
62. Johnson, P. C. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods Ecol. Evol.* **5**, 944–946 (2014).
63. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
64. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front. Neuroinform.* **7**, 14 (2013).
65. Moran, R. Optimal decision making in heterogeneous and biased environments. *Psychon. Bull. Rev.* **22**, 38–53 (2015).
66. Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W. & Forstmann, B. U. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J. Neurosci.* **32**, 2335–2343 (2012).
67. Decker, A., Dubois, M., Duncan, K. & Finn, A. S. Pay attention and you might miss it: greater learning during attentional lapses. *Psychon. Bull. Rev.* **30**, 1041–1052 (2023).
68. Seli, P. et al. On the clock: evidence for the rapid and strategic modulation of mind wandering. *Psychol. Sci.* **29**, 1247–1256 (2018).
69. Sarter, M., Givens, B. & Bruno, J. P. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Res. Brain Res. Rev.* **35**, 146–160 (2001).
70. Smallwood, J. Mind-wandering while reading: attentional decoupling, mindless reading and the cascade model of inattention. *Lang. Linguist. Compass*. https://doi.org/10.1111/j.1749-818X.2010.00263.x (2011).
71. Thomson, D. R., Besner, D. & Smilek, D. A resource-control account of sustained attention: evidence from mind-wandering and vigilance paradigms. *Perspect. Psychol. Sci.* **10**, 82–96 (2015).
72. Dong, H. W., Mills, C., Knight, R. T. & Kam, J. W. Y. Detection of mind wandering using EEG: within and across individuals. *PLoS One* **16**, e0251490 (2021).
73. Jana, S. & Aron, A. R. Mind wandering impedes response inhibition by affecting the triggering of the inhibitory process. *Psychol. Sci.* **33**, 1068–1085 (2022).
74. Jin, C. Y., Borst, J. P. & van Vugt, M. K. Predicting task-general mind-wandering with EEG. *Cogn. Affect. Behav. Neurosci.* **19**, 1059–1073 (2019).
75. Ratcliff, R. & Rouder, J. N. Modeling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356 (1998).
76. Evans, N. J. & Wagenmakers, E.-J. Evidence accumulation models: current limitations and future directions. *Quant. Methods Psychol.* **16**, 73–90 (2020).
77. Kucharský, Š, Tran, N.-H., Veldkamp, K., Raijmakers, M. & Visser, I. Hidden Markov models of evidence accumulation in speeded decision tasks. *Comput. Brain Behav.* **4**, 416–441 (2021).
78. Gupta, D., DePasquale, B., Kopec, C. D. & Brody, C. D. Trial-history biases in evidence accumulation can give rise to apparent lapses in decision-making. *Nat. Commun.* **15**, 1–15 (2024).
79. Aguillon-Rodriguez, V. et al. Standardized and reproducible measurement of decision-making in mice. *Elife* **10**, e63711 (2021).
80. Seli, P. et al. Mind-wandering as a natural kind: a family-resemblances view. *Trends Cogn. Sci.* **22**, 479–490 (2018).
81. Seli, P. et al. Intentionality and meta-awareness of mind wandering: are they one and the same, or distinct dimensions? *Psychon. Bull. Rev.* **24**, 1808–1818 (2017).
82. Smallwood, J., McSpadden, M. & Schooler, J. W. When attention matters: the curious incident of the wandering mind. *Mem. Cogn.* **36**, 1144–1150 (2008).
83. Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci.* **106**, 8719–8724 (2009).
84. Baer, R. A. Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin. Psychol.: Sci. Pract.* **10**, 125–143 (2003).
85. Schmertz, S. K., Anderson, P. L. & Robins, D. L. The relation between self-report mindfulness and performance on tasks of sustained attention. *J. Psychopathol. Behav. Assess.* **31**, 60–66 (2009).
86. Petranker, R. & Eastwood, J. D. Sitting with it: an investigation of the relationship between trait mindfulness and sustained attention. *Conscious. Cogn.* **90**, 103101 (2021).
87. Valentine, E. R. & Sweet, P. L. G. Meditation and attention: a comparison of the effects of concentrative and mindfulness meditation on sustained attention. *Ment. Health Relig. Cult.* **2**, 59–70 (1999).
88. Jha, A. P., Krompinger, J. & Baime, M. J. Mindfulness training modifies subsystems of attention. *Cogn. Affect. Behav. Neurosci.* **7**, 109–119 (2007).
89. Tang, Y.-Y., Hölzel, B. K. & Posner, M. I. The neuroscience of mindfulness meditation. *Nat. Rev. Neurosci.* **16**, 213–225 (2015).
90. Raichle, M. E. et al. A default mode of brain function. *Proc. Natl. Acad. Sci. Usa.* **98**, 676–682 (2001).
91. Poerio, G. L. et al. The role of the default mode network in component processes underlying the wandering mind. *Soc. Cogn. Affect. Neurosci.* **12**, 1047–1062 (2017).
92. Kajimura, S., Kochiyama, T., Nakai, R., Abe, N. & Nomura, M. Causal relationship between effective connectivity within the default mode network and mind-wandering regulation and facilitation. *Neuroimage* **133**, 21–30 (2016).
93. Brewer, J. A. et al. Meditation experience is associated with differences in default mode network activity and connectivity. *Proc. Natl. Acad. Sci. USA* **108**, 20254–20259 (2011).
94. Bonnelle, V. et al. Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. *J. Neurosci.* **31**, 13442–13451 (2011).
95. Mason, M. F. et al. Wandering minds: the default network and stimulus-independent thought. *Science* **315**, 393–395 (2007).

96. Scheibner, H. J., Bogler, C., Gleich, T., Haynes, J.-D. & Bermpohl, F. Internal and external attention and the default mode network. *Neuroimage* **148**, 381–389 (2017).
97. Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T. & Schwaiger, J. Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* **244**, 73–76 (1998).
98. Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P. & Lin, C.-T. Sustained attention in real classroom settings: an EEG study. *Front. Hum. Neurosci.* **11**, 388 (2017).
99. Wang, Y.-K., Jung, T.-P. & Lin, C.-T. EEG-based attention tracking during distracted driving. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**, 1085–1094 (2015).
100. Desantis, A., Chan-Hon-Tong, A., Collins, T., Hogendoorn, H. & Cavanagh, P. Decoding the temporal dynamics of covert spatial attention using multivariate EEG analysis: contributions of raw amplitude and alpha power. *Front. Hum. Neurosci.* **14**, 570419 (2020).
101. Bae, G.-Y. & Luck, S. J. Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *Neuroimage* **184**, 242–255 (2019).
102. Murray, S. & Krasich, K. Can the mind wander intentionally? *Mind Lang.* **37**, 432–443 (2022).
103. Maillet, D., Seli, P. & Schacter, D. L. Mind-wandering and task stimuli: stimulus-dependent thoughts influence performance on memory tasks and are more often past- versus future-oriented. *Conscious. Cogn.* **52**, 55–67 (2017).
104. Cole, S. N. & Tubbs, P. M. C. Predictors of obsessive-compulsive symptomology: mind wandering about the past and future. *Psychol. Res.* **86**, 1518–1534 (2022).
105. Unsworth, N. & McMillan, B. D. Similarities and differences between mind-wandering and external distraction: a latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychol.* **150**, 14–25 (2014).
106. Godwin, C. A., Smith, D. M. & Schumacher, E. H. Beyond mind wandering: performance variability and neural activity during off-task thought and other attention lapses. *Conscious. Cogn.* **108**, 103459 (2023).

Acknowledgements

We thank Todd Braver and Julie Bugg for helpful conversations. We would also like to thank the members of the Control and Decision Making Lab, the Control with Connections group, and the CCN-FACTS MURI group for advice and assistance. We are grateful to Leyla Vilic and Xintian Wang for help with data collection. This work was supported by a Multi-University Research Initiative grant (ONR/DoD N00014-23-1-2792) to W.K. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

C.Z. and W.K. designed the experiment; C.Z. developed the experiment and collected data; C.Z. analyzed data; C.Z. and W.K. 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/s44271-026-00424-9.

Correspondence and requests for materials should be addressed to Cathy Zhang.

Peer review information *Communications Psychology* thanks Matthew K. Robison, Ingmar Visser, Marta Migó for their contribution to the peer review of this work. Primary Handling Editors: Troby Ka-Yan Lui. A peer review file is available.