Distinct Electrophysiological Signatures of Intentional and Unintentional Mind-Wandering Revealed by Low-Frequency EEG Markers

Adrien Martel (ammmartel@gmail.com)
Institut du Cerveau (Paris Brain Institute)  https://orcid.org/0000-0002-6245-4206
Nicolas Bruno
Institut du Cerveau (Paris Brain Institute)
Paul Dockree
Ian Robertson
Jacobo Sitt
https://orcid.org/0000-0002-3878-4846
Antoni Valero-Cabre

Article

Keywords:

Posted Date: May 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2946669/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is NO Competing Interest.
Distinct Electrophysiological Signatures of Intentional and Unintentional Mind-Wandering Revealed by Low-Frequency EEG Markers

Abstract

Mind-wandering, a widespread mental phenomenon in which attention shifts from an ongoing task, varies in intentionality. Recent research highlights the importance of differentiating between intentional and unintentional mind-wandering, as the latter is uniquely linked to adverse outcomes such as psychopathologies. In this study, we explored the electrophysiological underpinnings of these distinct forms of mind-wandering using a robust feature extraction tool adapted from research on neural signatures of consciousness. We conducted univariate and multivariate pattern analyses on 54 EEG markers obtained from recordings immediately before participants provided multidimensional reports of their thoughts during a sustained attention task. Our findings revealed distinct electrophysiological signatures for on-task, intentional, and unintentional mind-wandering states, particularly within the low-frequency spectrum. Specifically, normalized theta power demonstrated the highest discriminative power for discerning on- and off-task states, while alpha band features and theta permutation entropy were uniquely associated with intentional versus unintentional mind-wandering. These results challenge the prevailing notion that increased alpha band activity is a generic marker of mind-wandering and suggest that unique brain activity patterns underlie the various forms of mind-wandering. Our study lays the groundwork for developing reliable, real-time detection systems for identifying mind-wandering using EEG machine learning models in both clinical and practical settings.

Introduction

Mind-wandering, a prevalent cognitive phenomenon, involves spontaneous thoughts unrelated to one's current activity or immediate environment (1). Estimated to occupy between 20% and 50% of waking mental activity (2), mind-wandering is associated with various functional outcomes in both the laboratory and daily life (3). Although mind-wandering can foster creativity, planning, and problem-solving (4, 5), uncontrolled occurrences can negatively impact performance across tasks (1) causing problems in educational (6), occupational (7) and operational settings (8). Given its ubiquity and link to a wide array of functional outcomes, understanding the nature of mind-wandering as a cognitive state, its neural basis and causal profile has emerged as a central goal within cognitive and clinical neuroscience (9–11).

Although the definition of mind-wandering is debated (10, 11), this work considers it a hypernym that encompasses a diverse range of experiences, and as task-unrelated thought (TUT) within the context of an explicit task. Investigating mind-wandering, an intrinsically covert state with few overt behavioral markers, poses unique challenges for both scientific inquiry and real-time mitigation. Researchers have relied on participants' introspective abilities, using 'experience
sampling’ methods with varying granularity, from binary on/off-task reporting to more detailed inquiries (for a review, see Weinstein et al. (12)). This research has illuminated how mind-wandering experiences differ across numerous phenomenological and cognitive dimensions (11).

A clinically significant distinction in mind-wandering is the presence or absence of intention. While some argue that TUT primarily results from unintentional failures of executive control (13), evidence suggests that mind-wandering can and does occur with intention (14). Intentionality refers to the extent that mind-wandering results from voluntary attention shifts towards TUT rather than dwindling externally directed attention (15). The intentionality dimension has emerged as a crucial factor with strong explanatory power, supported by its associations with specific content, neural, behavioral, and clinical correlates (14, 16–18).

Researchers employ experimental paradigms such as the Sustained Attention to Response Task (SART) (19) and experience-sampling methods like online thought-probes (20) to investigate and potentially manipulate mind-wandering. These thought-probes, considered the "gold-standard" of experience-sampling, intermittently interrupt tasks and prompt participants to classify their thoughts. Despite being the most effective method for assessing covert mental states, its validity faces challenges due to individual introspective ability variability and the influence of personal, contextual, and motivational biases (21–23). Additional concerns include modest-to-weak correlations between TUT rates in laboratory tasks and daily-life (7, 24) and the overuse of dichotomous thought-probes (12), which may inflate off-task reports (2).

To address these limitations, researchers have refined subjective measures by leveraging expert meditators' metacognitive abilities, expanding thought-probe categories (25), incorporating confidence ratings (23), and optimizing thought-probe frequency (26). Nevertheless, thought sampling requires task interruptions, which can increase participants' meta-awareness of their thought content, altering TUT reports and rates (27). A promising approach involves triangulating self-reports with objective measures like errors and reduced response variability (9), for a more comprehensive TUT assessment (21). Numerous studies have explored behavioral and physiological correlates of mind-wandering, assuming objective measures as ideal cognitive process markers. Prior research has identified behavioral markers like response time (13), task-related measures such as driving performance (8), and physiological measures, including skin conductance (28) and pupillometry (29). Although these objective measures show potential for complementing or replacing self-reports, they remain context-dependent, require specialized equipment, and provide limited insight into the brain dynamics underlying TUT.

Neural measures offer direct insights into the neurocognitive processes of covert mental states like mind-wandering. By leveraging robust markers and electromagnetic measures' high temporal resolution, real-time mental state detection becomes possible, creating opportunities for online mind-wandering mitigation (30) and reducing dependence on subjective methods.

The identification of electrophysiological correlates of mind-wandering has been a long-standing goal in the field, and early scalp EEG studies have primarily focused on event-related potentials (ERPs). Early potentials like P1 and N1, reflecting sensory stimulus-evoked responses, have been found to decrease during TUT episodes (30–34). The P3 amplitude, an index of general cognitive processing or attentional resource allocation (35), has also been observed to reduce during TUT in various studies (8, 29, 30, 32, 36, 37). Consistent with the perceptual decoupling hypothesis of mind-wandering (9), these findings suggest a strong link between a general attenuation in cortical
processing of external stimuli, short-term performance decline \((3, 38)\) and reduced overall vigilance \((39)\).

Beyond ERP activity, numerous studies have investigated the relationship between mind-wandering and oscillatory EEG activity. Time-frequency decomposed signal analyses within canonical frequency bands (i.e., delta (1-4Hz), theta (4-8Hz), alpha (8-14Hz), beta (15-30Hz), and gamma (30-50Hz)) have produced more variable results than ERP studies (for a review, see \((40)\)). There is substantial evidence supporting the critical role of alpha oscillations in gating brain information flow \((40, 41)\) and serving as a top-down inhibitory control mechanism to maintain task performance by suppressing task-irrelevant information \((42)\). Numerous studies have reported increased alpha power over frontal, central, parietal, and occipital scalp areas during TUT periods \((8, 29, 34, 37, 43-47)\), positioning alpha oscillations as a promising EEG signature of mind-wandering, despite some contrary evidence \((31, 39)\).

Greater alpha-band activity, along with attenuated ERP amplitudes, suggests reduced cortical processing of the external environment as attention shifts internally during TUT. This aligns with the executive function model of mind-wandering \((9)\), proposing that executive resources must decouple from sensory input during mind-wandering to protect internal processes from interference and enable TUTs to unfold uninterrupted \((48)\). Our study aims to expand on the limited research examining EEG differences between intentional and unintentional TUT, exploring electrophysiological markers that can reliably distinguish these attentional states. Previous findings have shown that ERPs, particularly the P3 component, are diminished during off-task states compared to on-task states, while greater alpha activity is associated with more intentional forms of TUT \((34, 49)\). Although increased alpha oscillations during TUT are relatively consistent, theta band activity patterns have exhibited more variability. Some studies observed greater theta activity during TUT compared to on-task states \((34, 44, 50, 51)\), while others reported the opposite pattern \((52, 53)\). Ample evidence suggests that theta rhythms within and across brain regions support executive control \((54)\) and attentional functions \((55)\). Some studies have reported increased frontal theta power when participants performed tasks imposing demands on externally oriented attention \((56)\).

Electrophysiological study findings collectively suggest the existence of a reliable EEG signature for TUT that can be effectively utilized using machine learning techniques. Recent studies have demonstrated the potential to predict TUT occurrence using EEG measures \((33, 43, 50, 57)\). Various classification approaches have been applied to diverse EEG features, with P3 and alpha \((29, 33, 50)\) emerging as the most characteristic of TUT, supporting findings from electrophysiological studies.

The aim of our study is to distinguish EEG signatures of intentional and unintentional TUT and investigate how they differ from on- and off-task states. We evaluated the discriminative ability of a wide range of putative EEG markers by performing a large-scale analysis on various predefined EEG features, testing their individual and collective ability to discriminate between on- and off-task activity, as well as iTUT and uTUT. This approach has been developed and successfully employed to robustly extract electrophysiological markers of consciousness across contexts and protocols, identifying features in the alpha and theta bands as indexes of consciousness \((58, 59)\). Given the close links between mind-wandering states, conscious content \((9, 60)\) and conscious access \((37)\) we extend this feature-extraction tool to TUT.
Applying this approach to an EEG dataset collected during a standard sustained attention test, we present the first systematic investigation of discriminative features for intentional and unintentional TUT. We anticipate identifying ERP markers, such as P3, as the most discriminative between on- and off-task states, and alpha-band markers to differentiate between intentional and unintentional TUT. By determining the electrophysiological signatures specific to different TUT states, we hope to contribute to a deeper understanding of the neural underpinnings of mind-wandering and its various subtypes.

**Results**

The analysis included 306 probe-caught and 935 self-caught reports, with 94 categorized as on-task, 82 as off-task, 286 as iTUT, and 250 as uTUT (refer to Table 1 for details).

**On- vs off-task**

Univariate Analysis of on-/off-task:

The results of the univariate analysis showed that the normalized power of the theta frequency band was the most discriminative marker for the on-/off-task contrast, with both the average (\(|\theta|_{\text{mean}}\)) and standard deviation (\(|\theta|_{\text{std}}\)) reaching statistical significance (AUC = 0.612, \(p_{\text{uncorrected}} = 0.01\) and AUC = 0.600, \(p_{\text{uncorrected}} = 0.023\), respectively) (see Fig. 1A). Nonetheless, after correcting for multiple comparisons with the false discovery rate (FDR) method, none of the measures for this contrast retained significance. Although effect sizes (i.e., AUC values above/below 0.50) suggested discriminative values, it should be noted that the current findings might benefit from further validation in studies with larger sample sizes. For more detailed results, refer to Table S1.

Multivariate Pattern Analysis (MVPA) of on-/off-task:

To assess the collective discriminative ability of the EEG markers for the on-/off-task contrast, an Extra Trees classifier was trained using all the markers identified in the previous univariate analysis. The accuracy of the classifier, computed as the mean of the 5-fold cross-validation, showed that the classifier performed above chance level with an AUC M of 0.603. We further assessed the statistical significance of this model using a 1000 permutation test, which demonstrated that the accuracy obtained was significantly above chance level (\(p = 0.042\); see Fig. 1C). Additionally, we obtained the feature importance scores of this classifier as a measure of univariate discriminative ability for each marker (see Fig. 1D). In line with the univariate analysis findings, both the average and standard deviation of the theta band power were identified as some of the most important features (see Fig. S1).

Contrasting on-task states with the other states:

To investigate whether the significant markers for on-/off-task were specific to this comparison or resulted from idiosyncrasies of the on-task condition, we conducted additional analyses. Specifically, we contrasted on-task against the other four categories of thought (i.e., about-task, distracted, iTUT, and uTUT) using the Synthetic Minority Over-Sampling Technique (SMOTE) to balance the samples across classes. Synthetic values are generated by SMOTE through linear interpolations of nearest-neighbor values, and it has been previously used for TUT classification due to the typical presence of unbalanced samples in mind-wandering experiments (33).
We repeated the same univariate procedure described in the previous section for each of
the contrasts (on-task/about-task, on-task/distracted, on-task/iTUT, and on-task/uTUT). The
analysis revealed that normalized theta power was the only significant marker against all
unfocused states conditions (distracted, iTUT, and uTUT; see Fig. 1B) and was consistently
higher for on-task than for any other category of thought. Moreover, the variation measure of this
marker reached significance after FDR correction for the contrasts of on-task against distracted
and iTUT. Additionally, normalized beta power was found to be higher for all the unfocused
states (distracted, iTUT, and uTUT) compared to on-task but was only significant for uTUT.
Notably, after multiple comparison correction, the ERP component P1 was found to be significant
and higher values were exhibited for on-task compared to uTUT. The only contrast for which no
markers were significant was between on- and about-task, possibly due to the similarity between
these two categories.

**Intentional vs Unintentional TUT**

Univariate Analysis of iTUT/uTUT:

In the univariate analysis of the iTUT/uTUT contrast, permutation entropy in the theta
frequency range averaged across trials ($\text{PE}_\theta\text{mean}: AUC = 0.625, p_{FDR} < 0.001$) was identified as the
most discriminative marker, with higher values for iTUT compared to uTUT (see Fig. 2A). The
normalized ($|\alpha|\text{mean}: AUC = 0.593, p_{FDR} = 0.002$) and non-normalized ($\alpha\text{mean}: AUC = 0.598, p_{FDR} =
0.004$) mean alpha power were also significantly higher for iTUT than uTUT. The most
discriminative marker for uTUT when compared to iTUT was increased normalized delta power
($|\delta|\text{mean}: AUC = 0.411, p_{FDR} = 0.005$). Additionally, unnormalized beta power ($\beta\text{mean}: AUC = 0.576,$
$p_{FDR} = 0.026$) and permutation entropy ($\text{PE} \beta\text{mean}: AUC = 0.431, p_{corrected} = 0.006$) were identified
as significant markers for classifying between iTUT and uTUT. For more detailed results, refer to
Table S2.

We repeated the same univariate analysis to determine if any marker could separate the three
unfocused categories (i.e., distracted, iTUT, and uTUT). None of the significant markers were
shared across the three contrasts (see Fig. S2). However, most of the markers that were found to be
significant for the contrast between iTUT and uTUT were also significant for the iTUT and the
distracted condition (e.g., $\text{PE} \theta\text{mean}, \text{PE} \beta\text{mean}, \text{PE} \gamma\text{mean}, |\alpha|\text{mean}, \alpha\text{mean}, |\delta|\text{mean} \text{ and } \beta\text{mean}$). Furthermore,
after FDR correction, no marker was found to be significant for the contrast of uTUT and the
distracted condition.

Comparison of univariate results for on/off-task and iTUT/uTUT:

A polar plot mapping the two sets of significant measures for the on-/off-task and
iTUT/uTUT contrast (Fig. 2B) allowed us to explore similarities in discriminative markers across
both contrasts. Two distinct groups of significant markers are displayed in the resulting figure,
indicating that unique EEG patterns characterize iTUT and uTUT.

To determine whether the computed average and standard deviation features were equally
informative in terms of their classification power across the two contrasts (on-/off-task and
iTUT/uTUT), AUC values for all markers were mirrored to obtain an absolute value of the
classifications and avoid direction bias. Then, a Mann Whitney U test was performed between the
mirrored AUC values of each marker for each contrast (i.e., the AUC of the on-/off-task contrast
against the AUC values for the iTUT/uTUT contrast) for the averaged and standard deviation
features. This analysis showed no significant differences for the markers based on average
calculations ($U = 248; p = 0.362$), However, significant differences were identified for the standard
deviation features ($U = 169; p = 0.018$) which were more discriminative for the on-/off-task comparison ($Median \text{ AUC} = 0.536$) than for the iTUT/uTUT contrast ($Median \text{ AUC} = 0.517$). For more details on this analysis, see Fig. S1.

MVPA of iTUT/uTUT:

The MVPA conducted on the iTUT/uTUT contrast yielded a slightly higher classification accuracy compared to the univariate analysis, with an AUC M of 0.632 in the cross-validation results. The permutation test further confirmed the model's classification performance was above chance ($p = 0.002$; see Fig. 2C). Additionally, we computed the feature importance from this model to provide additional insight into the discriminative power of each measure. Consistent with the univariate analysis, the results highlighted permutation entropy in the theta frequency band as the most important feature, followed by beta and alpha power (see Fig. 2D).

Discussion

Mind-wandering research has attracted substantial attention in psychology and neuroscience fields, given its widespread occurrence and strong links to detrimental functional consequences. However, the covert nature of mind-wandering poses a significant challenge to scientific inquiry, as the predominant data collection method, experience-sampling, is prone to biases. To address this issue, a growing number of studies are employing objective measures, like EEG and machine learning models, to identify attentional states. Research recognizes mind-wandering as a multifaceted construct that varies along numerous cognitive dimensions, with intentionality being a key predictor of functional outcomes (11, 25, 61). As a result, more fine-grained self-report measures are required to precisely encompass intentionality and study the complex processes underlying off-task thinking while addressing discrepancies in electrophysiological correlates across EEG research. Our research tackles this issue by implementing an extensive examination of numerous EEG features alongside multidimensional thought probes, offering a more precise depiction of the electrophysiological correlates of intentionality during mind-wandering.

Our study employed the SART and EEG recordings preceding multidimensional thought probe reports, which assessed if participants' thoughts were on-task, task-related, distracted, intentional, or unintentional task-unrelated thought (TUT). This allowed us to determine which of the 54 predefined markers, from four categories of EEG measures (ERP, spectral, connectivity, and permutation entropy), best characterized on- and off-task states, as well as intentional and unintentional TUT. Both univariate and multivariate analyses demonstrated that on-task states were reliably characterized by greater normalized power and variance in the theta range compared to off-task states, while intentional TUT (iTUT) was marked by increased theta permutation entropy and greater alpha power measures compared to unintentional TUT (uTUT). Specifically, the most discriminative EEG marker for on-task states was normalized power in the theta frequency range (4-7 Hz), as confirmed by an MVPA feature importance analysis and contrasting on-task states with other categories of thought. Similarly, the iTUT/uTUT contrast revealed that iTUT were associated with greater permutation entropy at the theta frequency range and increased alpha spectral measures. Notably, there were no overlapping discriminative markers across the contrasts, suggesting that on-task states, iTUT, and uTUT have distinct electrophysiological signatures. Our findings provide support for the notion that intentional and unintentional mind-wandering involve separate cortical architectures, with unique patterns of neural activity (14, 16, 34).
One of the key findings from our study was that while alpha EEG activity did not differentiate on- from off-task states, it did significantly discriminate between iTUT and uTUT (refer to Table 2). This challenges the prevailing notion that increased alpha band activity is a marker of mind-wandering and is different from most previous studies that have used tasks placing demands on externally oriented attention (8, 43–46, 53). However, studies that rely on tasks requiring internally oriented attention have reported reduced alpha activity during mind-wandering (39, 51, 62, 63), while increased alpha during tasks that do not require visual attention has been associated with better processing of internal states (64). In a recent study combining eye-tracking and EEG by Ceh et al. (47) a substantial correlation was found between occipital alpha power and pupillary diameter during intertrial rest periods. This suggests that both factors are probably connected to a shared gating mechanism that modulates sustained internal attention, with alpha increasing in response to internal attention requirements. Our results are consistent with this view, with intentional off-task thought characterized by increased alpha measures, a pattern often observed when individuals are engaged in internal tasks (41, 65–68). Furthermore, our findings are in line with the hypothesis that iTUT and uTUT reflect differences in the role that top-down processes play in ongoing thought. Specifically, we speculate that iTUT may be linked to the deliberate allocation of attention to internal processes, requiring a neurophysiological mechanism for top-down inhibition of irrelevant sensory input that is facilitated by alpha.

Conversely, more unintentional TUTs are likely to result from intermittent failures to maintain attention on the task. This is supported by the fact that discriminative markers in the iTUT/uTUT contrast were similarly discriminative for the contrast between iTUT and the 'distracted' category of thought, while the contrast between uTUT and the 'distracted' category did not yield any significantly discriminative markers. Participants were instructed to report being distracted when they attended to the external environment besides the task, and both states do not require active inhibition of sensory input. We speculate that the markers characterizing iTUT, such as alpha, reflect the application of control to organize an internal thought. Regarding markers characteristic of uTUT, delta normalized power was the most discriminative for this category compared to iTUT. Few studies have reported significant attentional state differences in the delta frequency range (for a review, see (69)), with one study linking decreased frontal delta activity to off-task states when compared to on-task during SART (53). Notably, a recent study reported that frontal delta activity predicted mind-wandering episodes (70). Since slow cortical rhythms are typically characteristic of sleep, the increase of delta features we observed during uTUTs may be indicative of a reduced alertness state. We also observed higher values of permutation entropy in the theta band as being discriminative of iTUT. Permutation entropy is an information complexity measure for time-series (94), which indexes the degree of conscious awareness in controls compared to anesthetized and minimally conscious state patients (71). Relatedly, Chen et al. (57) used several different classifiers (e.g., SVM, random forest, naive Bayes, and k-nearest neighbors) with standard spectral measures as well as spectral entropy measures and found that the random forest classifier performed best with entropy-based features. Increased measures of complexity, including permutation entropy, have been suggested to be necessary for a specific representation to be selected for conscious processing (59). Although we identified permutation entropy in the theta range as characteristic of iTUT, indicating increased conscious processing, the same measure of complexity in the beta and gamma range was also found to be characteristic of uTUT, demonstrating the relevance of such measures despite their challenging interpretation.

The second notable finding is related to the theta features that were characteristic of on-task states in contrast to off-task states. Specifically, our data indicate that periods reported as on-task are primarily marked by heightened normalized power at the theta frequency, in line with previous
research on executive functioning tasks and cognitive control (54, 72, 73). However, findings in the literature are inconsistent about theta activity across studies using the SART, with some reporting increases and others decreases of theta activity during TUT (for a review, see (69)). Furthermore, the small sample of on-task probe-caught reports in our study makes it difficult to draw any conclusive inference on the role of theta within the context of mind-wandering. Indeed, no discriminative markers for the contrast between on- and off-task states remained significant after correcting for multiple comparisons, likely due to the small ratio of probe-caught reports to EEG markers.

In contrast, the iTUT/uTUT contrast yielded multiple significant markers that survived correction, even though their discriminative power was comparable. Specifically, our data suggest that intentional TUTs (iTUTs) are associated with increased theta permutation entropy and greater alpha power measures compared to unintentional TUTs (uTUTs). Indeed, none of the discriminative markers for the contrast between on- and off-task survived correction for multiple comparisons, most likely due to the low ratio of probe-caught reports to EEG markers. Nevertheless, since the AUC calculation is less affected by sample size than p values, we maintain that the uncorrected significant markers presented here might still represent underlying electrophysiological distinctions.

In addition, our results demonstrate that oscillatory features exhibit high discriminative power, suggesting the potential for unobtrusive and continuous EEG monitoring without the need to elicit and measure a brain response as required for ERPs. Notably, our findings contradicted our initial predictions based on previous research, which found reduced amplitudes of P1 and P3 components to be significant predictors of mind-wandering (29, 33, 34).

Our study has certain limitations that must be addressed to provide a more accurate interpretation of the results. First, there is a risk of overfitting the models when using wide datasets with many features and a limited number of observations, which is the case for the on- and off-task contrast. However, despite the assumption that intentional and unintentional TUTs have more similarities than on- and off-task states, we obtained better classification performance for the former contrast. This result could be attributed to the much larger sample size available for intentional and unintentional TUT, reducing the likelihood of overfitting and providing more robust evidence for the distinct electrophysiological signatures of these attentional states. Second, the use of two types of thought probes used for both contrasts (probe-caught for the on-/off-task contrast and self-caught for the iTUT/uTUT contrast) may have contributed to some of the reported differences. Self-caught probes require participants to monitor their attention while performing the task, which could have introduced confounds. Third, while online thought sampling is a validated method to assess attentional states, potential biases may have introduced noise and impacted classification accuracy. However, until robust and task-independent markers of attentional states are identified, thought sampling will continue to be an indispensable tool. Additionally, the applicability of our results could be constrained, considering they rely on a specific task (SART) and emphasize certain attentional states.

Moreover, although we identified significant markers for the contrast between intentional and unintentional TUT, the discriminative markers for the on/off-task contrast did not survive correction for multiple comparisons, possibly due to the small sample size. Future studies should aim to collect more data points across different tasks to improve predictive performance and identify task-independent EEG markers. Multi-modal representation, combining EEG measures with indirect markers of mind-wandering, such as behavioral or pupillometric measures, could also...
be beneficial. Additionally, we should acknowledge that the SART may not be the most effective paradigm for studying the relationship between executive functions and mind-wandering. Future studies could explore this relationship using tasks that place higher demands on executive resources, such as the finger-tapping random generation task (74, 75). Lastly, investigating TUT modulation using brain stimulation constitutes an essential future avenue for establishing the causal importance of cortical areas in mind-wandering. This could include selectively downregulating maladaptive types of mind-wandering, such as unintentional TUT, and upregulating advantageous types, such as intentional TUT.

In summary, our study introduced a novel approach to investigating intentional and unintentional TUT by conducting comprehensive analyses on a broad range of EEG features. Our findings demonstrate the presence of unique electrophysiological signatures able to distinguish these two categories of thought from each other, with intentional TUT showing consistent differences from other attentional states. Specifically, we observed significant differences in EEG markers associated with top-down modulation of perception, such as increased alpha measures, suggesting that intentional TUT involves shielding internal processes by decoupling attention from sensory input. In conclusion, our study contributes to the growing body of evidence indicating that intentional and unintentional mind-wandering are conceptually distinct and supports the potential of EEG machine learning models for detecting TUT. By analyzing numerous EEG features and uncovering distinct electrophysiological signatures, our research paves the way for developing reliable, real-time detection systems for TUT, which can have promising applications in both clinical and practical settings. Ultimately, our work may lead to the replacement of subjective measures with objective measures, thus accelerating mind-wandering research and improving the development of interventions that can mitigate the negative consequences of unintentional mind-wandering. For example, practical applications of real-time TUT detection could include enhancing focus and productivity in work or educational settings by providing individuals with feedback on their attentional states. In clinical settings, such systems could be used to monitor and assess attentional deficits in individuals with ADHD or other attention-related disorders, informing the development of targeted interventions.

Materials and Methods

Dataset

Twenty-six participants (12 females, mean age: 25 years, standard deviation: 4.3 years) with normal or corrected-to-normal vision and no history of neurological or psychiatric disease, volunteered or received partial course credits to participate in the study. All procedures were approved by Trinity College Dublin ethics committee and conducted with adherence to the Declaration of Helsinki. Participants were informed extensively about the experiment, and all gave written consent.

Task stimuli and paradigm

Participants completed a fixed version of the sustained attention to response task (SART (19)), in a soundproof, electromagnetically shielded, dimly lit room. The SART is a computerized go/no-go task that requires participants to respond quickly to frequent targets (go: 1 to 5 and 7 to 9) and withhold responses to infrequent no-go targets (no-go: 6) presented amongst the background of targets. Digits from 1 to 9 were sequentially presented in the center of a monitor, with a presentation duration of 250 ms and an inter-stimulus interval (ISI) of 2316.5 ms (see Fig. 3). This
ISI was chosen to optimize the tradeoff between inducing a maximum number of mind-wandering episodes while not exacerbating the task’s difficulty by being too monotonous and imposing excessive demands on attentional resources. Stimuli were displayed at a font size of 140 in Arial font using the Presentation software package v19.0 (www.neurobs.com). Participants were instructed to press the left mouse button as quickly as possible in response to the target digit and to lock their response to the offset of the stimulus, a response strategy that has been shown to minimize both inter-individual variability in response times and speed-accuracy trade-offs (76).

The SART consisted of three blocks with a minimum duration of 8 min and a maximum duration of 15 min per block, resulting in a total task duration ranging from 24 to 45 min. The task included between 800 and 1200 trials approximately. The duration of each block was determined based on participants' self-reports, with each block lasting for at least 8 minutes or until three reports for each of the main categories of interest were obtained (on-task, TUTs). The average duration of the task was 38.7 minutes (SD: 4.8), and participants received an average of 65.3 thought probes (SD: 31.2), of which 71.4% were self-initiated. During the SART, participants were asked to report on their attentional state in two ways: through probe-caught and self-caught thought probes. Probe-caught probes were randomly and intermittently presented every 12, 18, 24, 30, 36, 42, or 48 trials, with an average frequency of every 30 trials (approximately every 28s, 42s, 56s, 76s, 83s, 97s, 111s, on average every 76s). Participants were prompted with the question "Where was your mind just now?" and asked to classify their ongoing thoughts before the interruption into five categories: 1) on task (focused attention), 2) about the task (task-related thoughts), 3) distracted (internal or external interference), 4) on future plans or memories (intentional TUT), and 5) daydreaming (unintentional TUT). Participants were also instructed to trigger a self-caught probe by pressing the space bar whenever they realized they were no longer on-task. While participants were told to report only being on-task during probe-caught reports, the 'on-task' option was available during self-caught probes to account for accidental button presses. Prior to the experiment, participants received training on the correct categorization of attentional states, which included examples for each category and a quiz to test their understanding. The categories for intentional and unintentional TUT were named 'on future plans or memories' and 'daydreaming', respectively, for convenience and clarity. To differentiate TUT from other types of mind-wandering, participants were instructed to report 'about task' when their thoughts related to their response strategy and 'distracted' when thoughts were related to internal sensations or external distractions such as noise in the environment.

EEG acquisition and preprocessing

The EEG was in line from 64 active channels placed on a cap according to the international 10-20 reference system, using the BioSemi Active Two system (www.biosemi.com). Continuous EEG data were amplified and digitized at 512 Hz, and bandpass filtered between 0.5 and 45 Hz. To assess eye movements and blinks, 4 electrooculography channels (EOGs) were used; two placed above and beneath the left eye and one on the outside of each eye. The preprocessing of EEG was performed with the MNE-python software package (77) (www.mne.tools). EEG data was down-sampled to 250 Hz before being high-pass filtered at 0.5 Hz and low-pass filtered at 45 Hz. Channels with excessively noisy signals were removed. Following our prior publication (59), the EEG data was segmented into epochs spanning -200 ms to 600 ms relative to SART stimuli and baseline corrected with respect to the pre-stimulus period (-200 to 0 ms). Noisy epochs were removed automatically with the Autoreject package (https://autoreject.github.io/). To correct for ocular and muscle artifacts, an ICA decomposition with the FastICA method was performed (78), and artifactual components were removed. Previously removed channels were interpolated using a spherical spline interpolation before re-referencing using a common average reference.
Additionally, the ERP component of all epochs was subtracted for the computation of all non-evoked markers. From this point on, only 5 epochs prior to a report (~12 s, 2.4 s per epoch, corresponding to 5 SART trials) were considered and labeled according to the category of thought reported by participants.

**Analysis**

Univariate and multivariate pattern analyses were conducted on all 54 markers across two contrasts. The first contrast compared probe-caught on- and off-task conditions with the off-task condition consisting of combined intentional and unintentional TUT epochs to balance sample size (see Martel et al. (34). The second contrast consisted of iTUT and uTUT reports from self-caught probes. For the univariate and multivariate analyses, we computed a total set of 27 markers drawn from Sitt et al. (59) and Engemann et al. (58), each belonging to one of four conceptual families: Event-Related Potentials (ERPs), Spectral, Information Theory, or Connectivity (see Table 3). Although Sitt et al. (59) employed a larger set of markers, we used a subset since many were found to be redundant and/or yielded poor performance. The selected set of markers mirrors the selection by Engemann et al. (58) with the addition of information theory and connectivity markers for all spectral bands, and excluding markers specific to their experimental methodology. For a detailed description and discussion of the markers, see Sitt et al. (59).

To aggregate the multiple observations per channel, per epoch, per time point, and/or per frequency bins, depending on the family of the marker, the observation points were aggregated first in the time/frequency/sensor domain, depending on the marker type, before being averaged across all sensors. Then, the last 5 epochs corresponding to one condition were aggregated via an average and the standard deviation, which is a measure of the variations between epochs (see Fig. 4). This yielded two different measures per epoch and marker, resulting in a total of 54 markers (27 average and 27 variation measures). The markers were labeled according to their type and to the final processing step, i.e., 'mean' or 'std' (e.g., $\alpha_{mean}$ or $\alpha_{std}$). The computation of each marker was performed using the NICE library (available at https://github.com/nice-tools/nice) for each of the 5 epochs immediately preceding a thought probe and linked to the condition corresponding to the self-report category. This time window of approximately 12 seconds preceding thought-probes is broadly consistent with previous analyses (28, 32, 39).

**Statistical analysis**

Univariate analysis

The univariate analysis of individual markers was conducted following the methodology described in Sitt et al. (59). The discriminative ability of a given marker was estimated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. The ROC curve displays the true positive rate (TPR) against the false positive rate (FPR) for different decision thresholds. An AUC value between 0.5 and 1 indicates a positive discriminative ability of a given feature for the first category (on-task or iTUT), while values between 0 and 0.5 indicate positive discriminative ability for the second category (off-task or uTUT). An AUC of 0.5 indicates chance levels of discrimination. The Mann-Whitney U test for independent samples was used to assess the statistical significance of discrimination for each marker. To account for multiple comparisons, the false discovery rate (FDR) method was used to correct for statistical significance. The FDR method controls the expected proportion of false positives among the significant results. We used an FDR threshold of 0.05 for determining statistical significance. These analyses were conducted for the entire set of 54 markers, comprising 27 averages and 27 standard deviations.
Multivariate pattern analysis (MVPA)

In line with the methods used in Engemann et al. (58), we employed an Extra-Trees classifier (79) from the Scikit-learn Python library (80) for the multivariate pattern analysis (MVPA). The Extra-Trees classifier is an ensemble learning method that builds multiple randomized decision trees as part of a forest to improve classification performance. It works by selecting random splits for each node in the tree, which reduces the computational cost and makes the model more robust to overfitting compared to traditional decision trees. Extra-Trees classifiers are non-parametric models that are robust and less sensitive to the scale of the input data, which makes them efficient at handling "wide" datasets containing more variables than observations. These models can also output feature importance scores that provide additional information on the discriminative capacity of individual features, including interdependency with other variables. We used 1000 trees with entropy as the impurity criterion, and other parameters set to default. The MVPA model was trained and tested using the same 54 markers used in the univariate analyses.

For cross-validation, we used a Monte Carlo cross-validation method with a training set size of 80% and a testing set of 20% with 5 iterations. To test the statistical significance of the MVPA model, we applied a 1000-permutation test for the same cross-validation procedure, resulting in 1000 samples with shuffled labels that were compared to the real cross-validation sample. The performance metric used to evaluate the Extra-Trees classifier was accuracy, which represents the proportion of correctly classified instances out of the total instances.

Supplementary Materials:
Fig. 1. Discrimination measures for the on/off-task contrast using AUC. AUC > 0.5 indicates higher measure for on-task than the other condition; AUC < 0.5 indicates the opposite pattern; AUC = 0.5 implies no discrimination. Filled circles represent \( p < 0.05 \) in a Mann-Whitney U test before multiple comparison correction; filled stars indicate significance after FDR correction. (A) AUC values of all markers for on-/off-task contrast, ranked by significant markers not surviving multiple comparison correction (filled circles) and non-significant markers (empty circles) in decreasing AUC order. (B) Scatter polar plot of AUC values for markers significant after correction (filled stars) in at least one comparison with on-task condition (about-task, distracted, iTUT, or uTUT). (C) Histogram of permutation tests evaluating statistical significance of MVPA for on-/off-task model, with histogram bins representing AUC of 1000 permuted models with shuffled labels; dashed line shows mean AUC for cross-validated models with correct labels. (D) Average feature importance measures from MVPA for top 10 features in on-/off-task model.
Fig. 2. Discrimination measures for iTUT/uTUT and on/off-task contrasts using AUC. AUC>0.5 indicates higher measure for the first condition (iTUT or on-task) than the second (uTUT or off-task); AUC<0.5 indicates the reverse pattern; AUC=0.5 implies no discrimination. Filled circles represent p<0.05 in a Mann-Whitney U test before multiple comparison correction; filled stars indicate significance after FDR correction. (A) AUC values of all markers for iTUT/uTUT comparison, ranked by significance after multiple comparison correction (filled stars), significant markers not surviving multiple comparison correction (filled circles), and non-significant markers (empty circles). (B) Scatter polar plot of AUC values for significant markers in on-/off-task or iTUT/uTUT contrasts. (C) Histogram of permutation tests evaluating statistical significance of MVPA for iTUT/uTUT model, with histogram bins representing AUC of 1000 models with shuffled labels; dashed line shows mean AUC for cross-validated models with correct labels. (D) Average feature importance measures from MVPA for top 10 features in iTUT/uTUT model.
Fig. 3. Sustained-Attention-to-Response Task (SART) featuring thought probes. Participants observed a continuous sequence of single digits, pressing a button for each digit except the number 6 (targets). Attentional state was assessed intermittently through probe-caught probes (occurring on average every 30 trials) or self-caught probes initiated by pressing the space bar.
Fig. 4. EEG feature extraction pipeline for univariate and multivariate analyses. EEG markers were categorized into four conceptual families: evoked responses, spectral, information theory, and connectivity. These were aggregated in three steps for the five SART trials preceding a probe response. (1) EEG features from ERP, spectral, or connectivity families were averaged across time, frequency, or sensor dimensions, respectively. (2) Markers were averaged over all EEG channels to yield one unique value per epoch. (3) Two features were extracted from each marker (indicated by red dots) by calculating both mean and standard deviation. Markers were labeled according to type and final computation step, e.g., standard deviation of alpha oscillations was labeled $\alpha_{\text{std}}$. Univariate analysis employed ROC-AUC metric for each of the 54 markers to assess individual classification performance for both contrasts (on-task/off-task and iTUT/uTUT). Multivariate analysis evaluated the collective discriminative ability of all markers for both contrasts using Extra Trees Classifier.

Table 1. Count of self-reports per thought category, separated by probe-caught and self-caught probes.

<table>
<thead>
<tr>
<th>Probe</th>
<th>On-task</th>
<th>About-task</th>
<th>Distraction</th>
<th>iTUT</th>
<th>uTUT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe-caught</td>
<td>94</td>
<td>82</td>
<td>48</td>
<td>39</td>
<td>43</td>
<td>306</td>
</tr>
<tr>
<td>Self-caught</td>
<td>-</td>
<td>200</td>
<td>199</td>
<td>286</td>
<td>250</td>
<td>935</td>
</tr>
</tbody>
</table>

Table 2. Overview of significant EEG markers obtained in univariate analyses for all three classifications and the MVPA accuracy for each of them. The markers are presented from the highest to the lowest AUC. For the probe caught (PC) vs self-caught (SC) probe classification, the results shown correspond to the under-sample method.

<table>
<thead>
<tr>
<th>Classification</th>
<th>On- vs. Off-Task</th>
<th>iTUT vs. uTUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR corrected markers</td>
<td>-</td>
<td>PE $\theta_{\text{mean}}$; $\alpha_{\text{mean}}$; $</td>
</tr>
<tr>
<td>Uncorrected markers</td>
<td>$</td>
<td>\theta</td>
</tr>
<tr>
<td>MVPA (AUC)</td>
<td>0.60</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 3. Description of the full list of EEG-markers used and the category to which they pertained.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Marker</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td>Contingent Negative Variation</td>
<td>ERP</td>
</tr>
<tr>
<td>P1</td>
<td>P100 evoked potential</td>
<td>ERP</td>
</tr>
<tr>
<td>P3a</td>
<td>P3a evoked potential</td>
<td>ERP</td>
</tr>
<tr>
<td>P3b</td>
<td>P3b evoked potential</td>
<td>ERP</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Alpha PSD</td>
<td>Spectral</td>
</tr>
<tr>
<td>$</td>
<td>\alpha</td>
<td>$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Beta PSD</td>
<td>Spectral</td>
</tr>
<tr>
<td>$</td>
<td>\beta</td>
<td>$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Delta PSD</td>
<td>Spectral</td>
</tr>
<tr>
<td>$</td>
<td>\delta</td>
<td>$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Gamma PSD</td>
<td>Spectral</td>
</tr>
<tr>
<td>$</td>
<td>\gamma</td>
<td>$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Theta PSD</td>
<td>Spectral</td>
</tr>
<tr>
<td>$</td>
<td>\theta</td>
<td>$</td>
</tr>
<tr>
<td>MSF</td>
<td>Median Power Frequency</td>
<td>Spectral</td>
</tr>
</tbody>
</table>
SE90  Spectral Edge 90  Spectral
SE95  Spectral Edge 95  Spectral
SE  Spectral Entropy  Spectral
K  Kolgomorov Complexity  Information Theory
PE α  Permutation Entropy Alpha  Information Theory
PE β  Permutation Entropy Beta  Information Theory
PE δ  Permutation Entropy Delta  Information Theory
PE γ  Permutation Entropy Gamma  Information Theory
PE θ  Permutation Entropy Theta  Information Theory
wSMI α  weighted Symbolic Mutual Information Alpha  Connectivity
wSMI β  weighted Symbolic Mutual Information Beta  Connectivity
wSMI δ  weighted Symbolic Mutual Information Delta  Connectivity
wSMI γ  weighted Symbolic Mutual Information Gamma  Connectivity
wSMIθ  weighted Symbolic Mutual Information Theta  Connectivity

References


