Introspection confidence predicts EEG decoding of self-generated thoughts and meta-awareness

Naya Polychroni[1, a], Maria Herrojo Ruiz[1,2], & Devin B. Terhune [1]

[1]*Department of Psychology, Goldsmiths, University of London, UK*

[2]*Center for Cognition and Decision Making, National Research University Higher School of Economics, Russian Federation*

[a] Corresponding author:

Naya Polychroni

Department of Psychology

Goldsmiths, University of London

8 Lewisham Way

New Cross, London, UK SE14 6NW

ppoly002@gold.ac.uk

Abstract

The neurophysiological bases of mind wandering (MW) – an experiential state wherein attention is disengaged from the external environment in favour of internal thoughts, and state meta-awareness are poorly understood. In parallel, the relationship between introspection confidence in experiential state judgements and neural representations remains unclear. Here, we recorded EEG whilst participants completed a listening task within which they made experiential state judgments and rated their confidence. Alpha power was reliably greater during MW episodes, with unaware MW further associated with greater delta and theta power. Multivariate pattern classification analysis revealed that MW, and meta-awareness can be decoded from the distribution of power in these three frequency bands. Critically, we show that individual decoding accuracies positively correlate with introspection confidence. Our results reaffirm the role of alpha oscillations in MW, implicate lower frequencies in meta-awareness, and are consistent with the proposal that introspection confidence indexes neurophysiological discriminability of representational states.

Keywords: EEG; confidence; meta-awareness; mind wandering; support vector machine

1 Introduction

Our brains are constantly bombarded by dynamic sensory input, yet we frequently shift away from the external environment towards thoughts, emotions and images that do not emerge from ongoing perceptual processes, are self-generated and unrelated to one's current task (*mind wandering*; Antrobus et al. 1970; Christoff et al. 2016; Smallwood and Schooler 2015). Research suggests that mind wandering (MW) occurs in ~30-50% of our waking hours [3,4], has deleterious effects on sensory and cognitive processing with corresponding reductions in event-related potentials (ERP) in response to external stimuli (Smallwood et al. 2008; Barron et al. 2011; Kam et al. 2011; Baird et al. 2014). These effects produce concomitant negative effects in a variety of tasks from driving [7] to reading [8–10].

Converging findings implicate default mode network (DMN; [11,12] in mind wandering [13] and its associated cognitive operations such as self-related processing [14], autobiographical memory, theory of mind, and future planning [15–17]. Mind wandering episodes can occur with (*tuning-out*) and without (*zoning-out*) meta-awareness (Schooler et al. 2004, 2011; Smallwood, McSpadden, et al. 2007) with the latter suggested to reflect a more pronounced form of mind wandering characterized by poorer performance (Smallwood, McSpadden, and Schooler 2007) and greater recruitment of DMN and executive control network [13].

Despite advances in its network architecture, the patterns of oscillatory activity underpinning mind wandering are poorly understood [20]. Multiple studies indicate that mind wandering states [21–24], and particularly zoning-out (unaware mind wandering) (Boudewyn and Carter 2018), are characterized by elevated alpha (~8-12Hz) power. Alpha oscillations are suggested to support inhibition-related processes [25,26], and attentional suppression [27] and are further implicated during working memory and mental imagery tasks [28], internally-oriented brain states [29,30], and inner speech [31], all of which figure prominently in the experience of mind wandering. However, at least two studies failed to replicate these effects [32,33] and observed greater delta (~2-3Hz) and theta (~4-7Hz) power during mind wandering. Slow-wave brain oscillations are typically associated with decreased sustained task-related attention [34]. Delta frequency contributions have also been shown during increased focus on internal processing and pertinent inhibition of interference [35] whereas theta activity has been consistently shown to relate to maintenance of information in working memory [34,36]. Discrepancies in the observed association between alpha and mind wandering are plausibly attributed to the tasks and methods used in the aforementioned studies [32,33]. That is, lower alpha activity during mind wandering episodes in some studies might be due to the concurrent task (breath-counting) involving internally-focused attention and counting (Palva et al. 2005; Sauseng et al. 2005). In parallel, it is difficult to compare findings between self-reports that are prompted (probe-caught) and the above studies due to the latter using self-caught measures (participants are asked to indicate when they catch themselves mind wandering) of mind wandering, which likely capture shifts from internal to external focus that probably involve different mechanisms to the occurrence of mind wandering.

Measuring the neurophysiology of self-generated thoughts requires experience sampling methods in which participants report on aspects of their experience, thereby affording a prominent role to introspective abilities in the assessment of mind wandering [39]. A neglected feature of these abilities within the context of mind wandering is confidence in these introspective reports. Emerging evidence suggests that confidence reflects variability in access to experiential states (Fleming & Lau 2014; Seli et al. 2015) and thus is likely to be highly informative in elucidating variability in the phenomenology and neurophysiology of mind wandering episodes. Confidence in perceptual judgements [40], positively correlates with decision accuracy [41–43], and reliably tracks ERP dynamics related to error [44] and sensory processing (e.g. Zakrzewski et al. 2019). Confidence in mind wandering reports has to date been neglected but preliminary work has shown that it varies greatly within and between individuals and moderates the relationship between response time variability and self-reports of mind wandering (Seli et al. 2015) (but see Meier 2018). Nevertheless, the neurophysiology of these effects is unknown.

One possibility is that if confidence reflects superior access to experiential states, high confidence would be associated with more clearly dissociable neural representations. Multivariate pattern classification analysis (MVPC) has been successfully used to identify the mapping between distributed patterns of neural activity and corresponding mental states (Haxby et al. 2001; Haynes and Rees 2006; Jin, Borst, and van Vugt 2019; Mittner et al. 2014). An advantage of MVPC allows researchers to assess whether shared information across multiple features (e.g., channels, frequency, time points) encodes class-related information (Haxby et al. 2001). Using these methods, recent research has revealed associations between decoding accuracy and individual differences in perceptual discrimination (Kim et al. 2015) and intra-individual variability in confidence [53]. The extent to which experiential states can be decoded, reflecting multivariate dissimilarity of neural representations, may thus underlie confidence in the corresponding mental representations.

The present study investigated the oscillatory dynamics of experiential states using an ecological task lacking performance indicators [54] in order to examine the neurophysiological basis of mind wandering, dissociate meta-awareness of mind wandering (henceforth state meta-awareness) and investigate the neurophysiological implications of participants' confidence in self-reports. During concurrent EEG recording, participants listened to an audiobook and were intermittently probed regarding their experiential state and state meta-awareness, and rated their confidence in both judgements. We expected that mind wandering would be characterized by elevated alpha power (Compton et al. 2019) whereas unaware mind wandering would additionally be associated with greater power in slow oscillatory bands (delta, theta) [32,33]. Motivated by previous research showing that joint activity patterns across different features (e.g. frequency bands) can be more informative of mental representations than univariate information (Allefeld and Haynes 2015), we then assessed whether distributed information across patterns of EEG spectral features could be used to decode different experiential states using MVPC. These analyses were further guided by our aim to evaluate the hypothesis that introspection confidence in experiential states reflects higher dissimilarity of the underlying neural representations.

2 Results

2.1 Characteristics of mind wandering, state meta-awareness and introspection confidence

During the audiobook listening task, participants (*N*=39) reported mind wandering (*M*%±*SD*) on 39.7±17.9 of the probes, with stable rates across blocks (block 1: 41.2±20.6; block 2: 40.9±21.6; block 3: 36.9±19.0). Among mind wandering states, participants reported tuning-out (56.6±19.0) more often than zoning-out (43.4±19.0). Participants varied (range, *M*%±*SD*) in their confidence for ES judgements (14.8-91.4, 61.7±7.7), and displayed less confidence in mind wandering reports (11.3-90.0, 53.2±19.9), than in on-task reports (8.6-94.2, 64.4±19.0), $t(38)=4.78$, $p<.001$, $g=.57$, [0.32 0.90]. Specifically, on-task reports were rated with significantly higher confidence than both tune-out ($t(38)=2.39$, $p=.02$, $g=.34$, [0.08 0.67]) and zone-out reports ($t(38)=3.80$, $p<.001$, $g=.58$, [0.28 0.96]). Participants were moderately confident in their MA judgments during mind wandering states (16.7-92.0, 58.1±6.3), with numerically, albeit non-significantly, greater confidence in tune-out (17.2-92.3, 58.0±18.8) than in zone-out (6.9-90.5, 52.3±21.8) reports, $t(38)=1.55$, $p=.13$, $g=.27$, [-0.06 0.64]. These rates are similar to previous research (Christoff et al. 2009; Seli et al. 2015; Varao Sousa et al. 2013) and demonstrate variability in experiential states, state meta-awareness, and introspection confidence during the task.

2.2 Audiobook listening assessment and mind wandering frequency

Accuracy on the assessment averaged across blocks (*M%*±*SD*: 74.5±10.0) and was above chance performance (50%, one sample *t*-test: $t(38)=15.3$, $p<.001$, $g=2.44$, [1.84, 3.59]). Performance was comparable across blocks (block 1: 72.3±12.0, block 2: 74.6±14.4, block 3: 76.5±12.1), suggesting stable motivation throughout the task. Mind wandering frequency reliably significantly correlated negatively with assessment accuracy in the last two blocks (block 1: r=-.29 [95% CI: -.53, .02], p=.079 (p=.017, with outliers), block 2: r_s=-.43 [-.68, -.10], p=.009 (p<.001, with outliers), block 3: r=-.41 [-.68, -.11], p=.011 (p=.011, with outliers), thereby providing an indirect behavioural validation of participants' self-reports and corroborating previous research (Schooler et al. 2004; Boudewyn and Carter 2018).

2.3 Oscillatory characteristics of mind wandering and state meta-awareness

As expected, the cluster-based permutation test revealed greater alpha power during mind wandering than on-task states (**Figure 1**). The analysis revealed two temporally-adjacent clusters just prior to probe onset, p=.004, g=0.56 [0.32, 0.92]; p=.024, g=0.50 [0.29, 0.80]. Both effects were topographically diffuse and most pronounced over bilateral frontocentral and right posterior sites. Similarly, we observed greater alpha power during tuning-out than on-task states in a single cluster, p=.008, g=0.66 [0.37, 1.05] (**Supplementary Figure 1**). This effect was also close to probe onset and was primarily observed over fronto-central and parieto-occipital regions. There were no other significant differences between states in the other frequency bands.

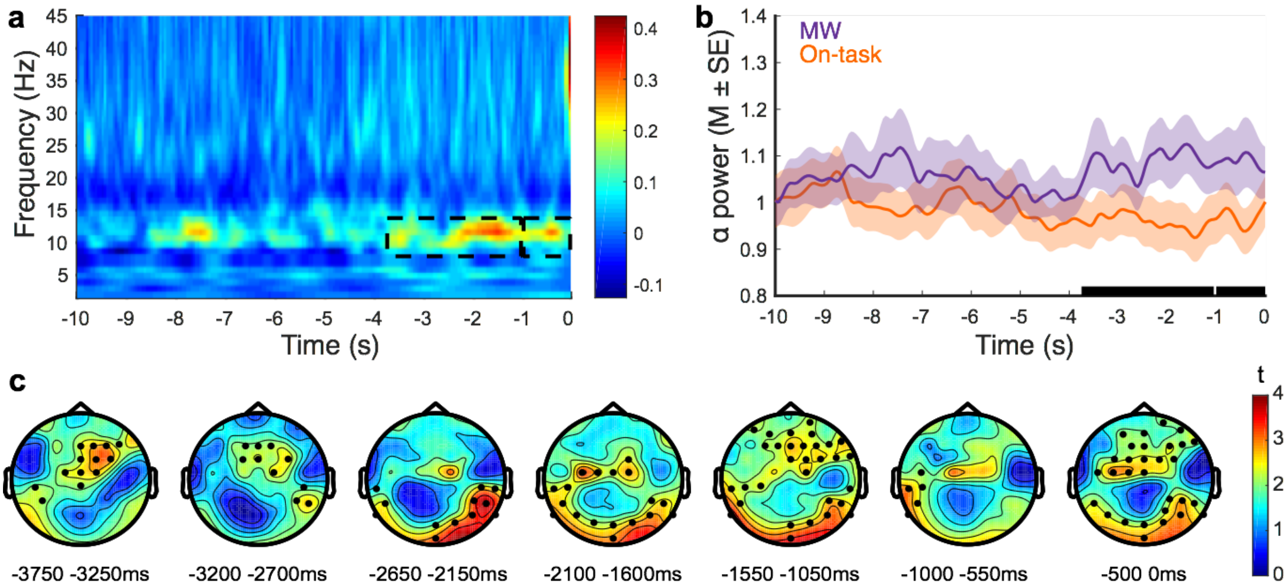


Figure 1. Oscillatory differences between states (MW – on-task, *N*=39) as a function of time relative to probe onset (0s). a) Time-frequency decomposition averaged across electrode sites. Broken black rectangles denote spectrotemporal clusters reflecting significant state differences ($p<.025$, two-sided cluster-based permutation test). b) Alpha (8-13Hz) spectral power averaged over the electrode sites of the two clusters (significance denoted by black bars on the x-axis). c) Topography of the clusters at different 500/550ms sub-windows (black markers denote electrodes that were present on at least 50% of samples in each time window). MW = Mind Wandering.

In line with the foregoing results, zoning-out (unaware mind wandering) states were characterized by greater power than on-task states in delta, theta, and alpha bands in two distinct time windows (**Figure 2**). Alpha power was greater for zoning-out than on-task states in a short interval early in the epoch, p=.004, g=0.75 [0.51, 1.06], and a long interval just prior to probe onset, p=.002, g=0.86 [0.57, 1.28]. Both effects were larger in magnitude than the comparisons between on-task and mind wandering and tune-out states and topographically diffuse but strongest over bilateral frontal and posterior sites. Similarly, theta power was greater in zoning-out than on-task states in a single window overlapping with the late alpha effect, p=.010, g=0.83 [0.54, 1.24]; this effect was larger in right frontal electrodes but over time shifted to left temporo-parietal sites. A second theta band cluster overlapped in time with the early alpha cluster but did not achieve significance despite a large effect size, p=.028, g=0.76 [0.50, 1.14]. Zoning-out states were also associated with greater delta power than on-task states in two clusters that were temporally coincident with the foregoing effects but with substantially larger effect sizes, p=.018, g=1.43 [1.04, 2.16]; p=.018, g=1.16 [0.74 1.73]. These effects were topographically more focal and largely restricted to midline central electrodes.

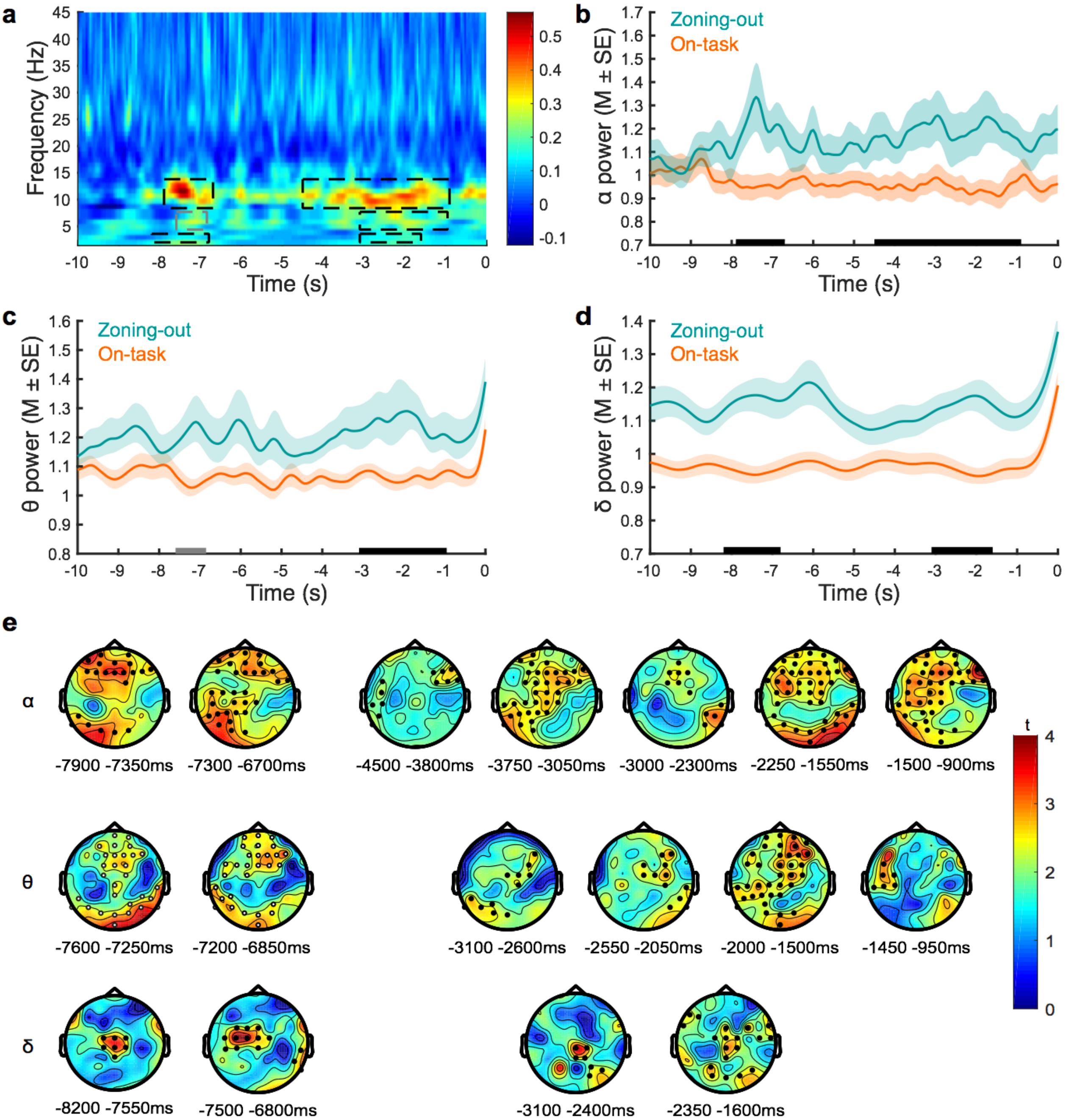


Figure 2. Oscillatory differences between states (zoning-out – on-task, N=25) as a function of time relative to probe onset (0s). a) Time-frequency decomposition averaged across electrode sites. Broken black rectangles indicate spectrotemporal clusters reflecting a significant difference (p<.025, two-sided cluster-based permutation test) and the grey rectangle indicates a trend-level (.025<p<.05) cluster. b, c, d) Alpha (8-13Hz), theta (4-7Hz) and delta (2-3Hz) spectral power averaged over the electrode sites of the clusters (black bars=significant, grey bar=trend). e) Topography of the clusters at different sub-windows within the cluster (black markers denote

electrodes that were present on at least 50% of samples in each time window, white electrodes mark topography of the trend-level effect).

Zoning-out states were also characterized by greater theta power than tuning-out states in a single cluster close to probe onset similar to previous effects at a trend-level of significance, p=.026, g=0.86 [0.45 1.41] (**Supplementary Figure 2**). This effect was most pronounced in parieto-central electrodes. Failure to reach significance is plausibly due to the reduced sample for this analysis (N=21) due to MW trial partitioning. There were no other significant effects in the other frequency bands.

2.4 Multivariate pattern classification analyses

The univariate analyses of the EEG signals managed to identify spectral and spatio-temporal differences between states. This analysis, however, failed to demonstrate a robust and significant difference between aware and unaware mind wandering. In the following, we present the results for two sets of analyses: the time-*varying* MVPC analyses performed in the epoch before probe onset (-10 to 0s; 21 bins of 500ms) and the time-*averaged* MVPC analyses performed after collapsing temporal information for decoding across the entire epoch.

2.4.1 Time-varying MVPC

For each two-class (state) comparison, we tested four models that used as features the spectral power in the 64 channels in the delta, theta and alpha frequency bands (multi-frequency model) and in these three bands separately (single-frequency band models). At the participant-level, we report the percentage of participants for which significant decoding was obtained in at least one bin.

The multi-frequency model (trials per class: 23.36±9.01) significantly decoded mind wandering from on-task states in 26% of participants, with similar proportions in single-frequency models: delta: 23%; theta: 15%; and alpha: 26% (**Supplementary Figure 3a)**. At the group level, both the multi-frequency and the single-frequency models were able to decode mind wandering and on-task states with significant accuracies observed at various bins in the window (**Figure 3a)**. Similarly, the multi-frequency model (trials per class: 18.78±8.20) significantly decoded on-task from tuning-out states in 22% of participants although the single-frequency models displayed superior decoding: delta: 44%; theta: 30%; alpha: 33% (**Supplementary Figure 3b)**. Group-level classification was significant in four bins in the multi-frequency model with similar results for theta and alpha, but not delta (**Figure 3b**). Significant decoding accuracy between zoning-out and on-task states (trials per class: 13.88±5.12**)** was observed in 24% of participants for the multi-frequency model, with comparable or superior accuracy in the single-frequency models (delta: 44%; theta: 36%; alpha: 24%; **Supplementary Figure 3c**). Group-level significant decoding was found for the multi-frequency model and the single-frequency models (**Figure 3c**). The theta and delta effects were spread throughout the window with the delta model exhibiting decoding accuracy across the full window. Tuning-out and zoning-out states (trials per class: 12.57±3.63) were decoded in 14% of participants in the multi-frequency model, with superior decoding in the single-frequency models (delta: 24%; theta: 19%; alpha: 29%, **Supplementary Figure 3d**). At the group-level, the multi-frequency and single-frequency models showed significant decoding although this was restricted to one bin in the former and generalized throughout the window in the latter (**Figure 3d**). These results indicate that decoding for all two-class comparisons was significant in several time bins for models using both the multi-frequency model and the reduced feature space (single-frequency) models suggesting that the individual frequency bands alone are sufficient to decode experiential states and state meta-awareness.

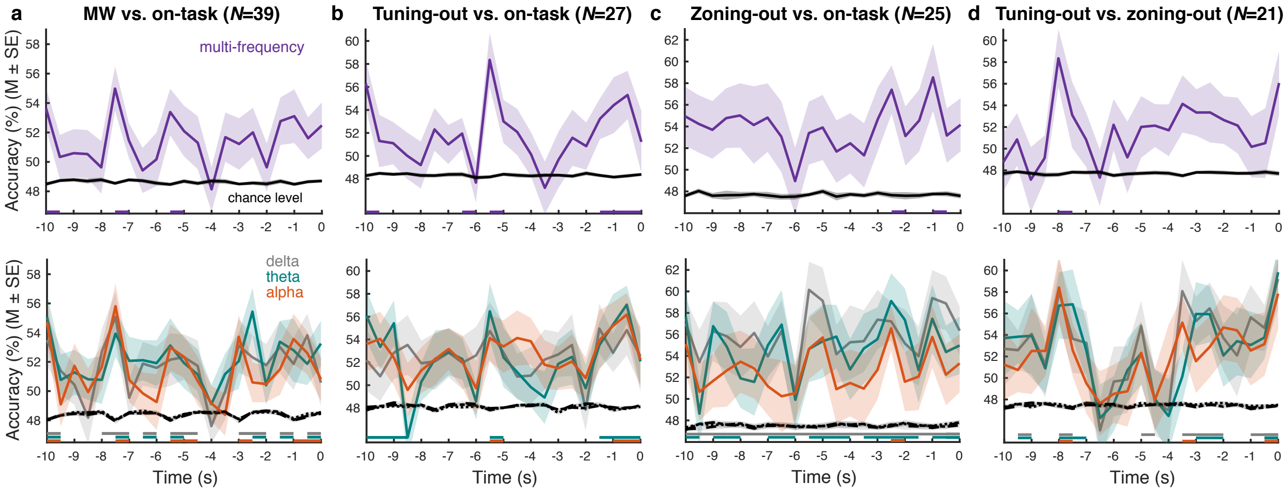


Figure 3. Time-varying MVPC (Multivariate Pattern Classification analysis): Group decoding accuracy across participants in multi-frequency (**top panels**) and single-frequency (**bottom panels**) models for four different state contrasts. The null distribution was obtained by 5000 random permutations after shuffling the labels (black lines). Above chance decoding (relative to the permutation distribution) is denoted by horizontal x-axis coloured markers (based on an FDR correction; see **Supplementary Table 1**). MW = Mind Wandering.

2.4.2 Time-averaged MVPC

For time-averaged MVPC, the multi-frequency model significantly decoded mind wandering from on-task states in 28% of participants (**Figure 4a**). Individual-frequency models decoded states in similar proportions of participants: delta: 21%; theta: 26%; alpha: 23% (**Supplementary Figure 4e**), and did not significantly differ, $p>pth$ (multi-frequency model vs. single-frequency models). At the group-level (**Supplementary Figure 4a**), the multi-frequency model displayed significant decoding accuracy (range, $M\%{\pm}SE$: 37.4-86.9, 57.25±1.99, p=0), as did single-frequency models: delta (22.22-83.33, 56.75±1.85, p<.001), theta (16.66-95.83, 56.68±2.35, p<.001), and alpha (35.55-81.35, 56.96±1.53, p=0).

The multi-frequency model decoded tuning-out from on-task states in 19% of participants (delta: 11%; theta: 15%; alpha: 19%, **Supplementary Figure 4f**), with no significant differences between models, $p>pth$. At the group-level (**Supplementary Figure 4b**), the multi-frequency model displayed significant decoding accuracy (36.66-87.5, 57.20±2.12, $p<.001$), as did the single-frequency models (delta: 28.33-86.66; 54.91±2.54, $p=.011$; theta: 33.33-95.83, 58.68±2.26, $p=0$; alpha: 36.11-79.16, 55.26±2.21, $p=.003$).

The multi-frequency model also significantly decoded zoning-out from on-task states in 44% of participants. The single-frequency models (**Supplementary Figure 4g**) significantly decoded these states in fewer participants (delta: 32%; theta: 28% [**Figure 4b**]; alpha: 32%), but these differences were not significant, $p>pth$. At the group-level (**Supplementary Figure 4c**), the multi-frequency model showed the greatest decoding accuracy (33.88-47.37, 63.70±2.79, $p=0$), albeit with comparable accuracies in the single-frequency models (delta: 34.44-90.28, 60.32±3.09, $p<.001$; theta: 15.28-94.44, 59.11±3.18, $p=.002$; alpha: 31.94-88.88, 61.39±2.66, $p<.001$).

Meta-awareness of mind wandering (zoning-out [unaware] vs. tuning-out [aware]) was significantly decoded in 29% of participants by the multi-frequency model (**Supplementary Figure 4h**). Performance varied across single-frequency models (delta: 19% [**Figure 4c**]; theta: 10%; alpha: 38%), but did not significantly differ, $p>pth$. At the group-level (**Supplementary Figure 4d**) the multi-frequency model displayed significant decoding (23.33-89.17, 55.50±3.43, $p=.027$), as did the alpha model (19.17-90.00, 59.50±3.88, $p=.005$), and delta model (34.72-100, 56.64±3.22, $p=.007$) but not the theta model (6.67-94.44, 53.35±3.76, $p=.154$). Owing to the fluctuating nature of mind wandering states, we expected that time-varying MVPC would reveal less consistent individual and group-level effects than time-averaged MVPC. However, the results did not suggest any substantial differences between the two approaches. Both performed similarly for all contrasts, albeit the time-averaged MVPC using theta frequency information was poor in decoding state meta-awareness.

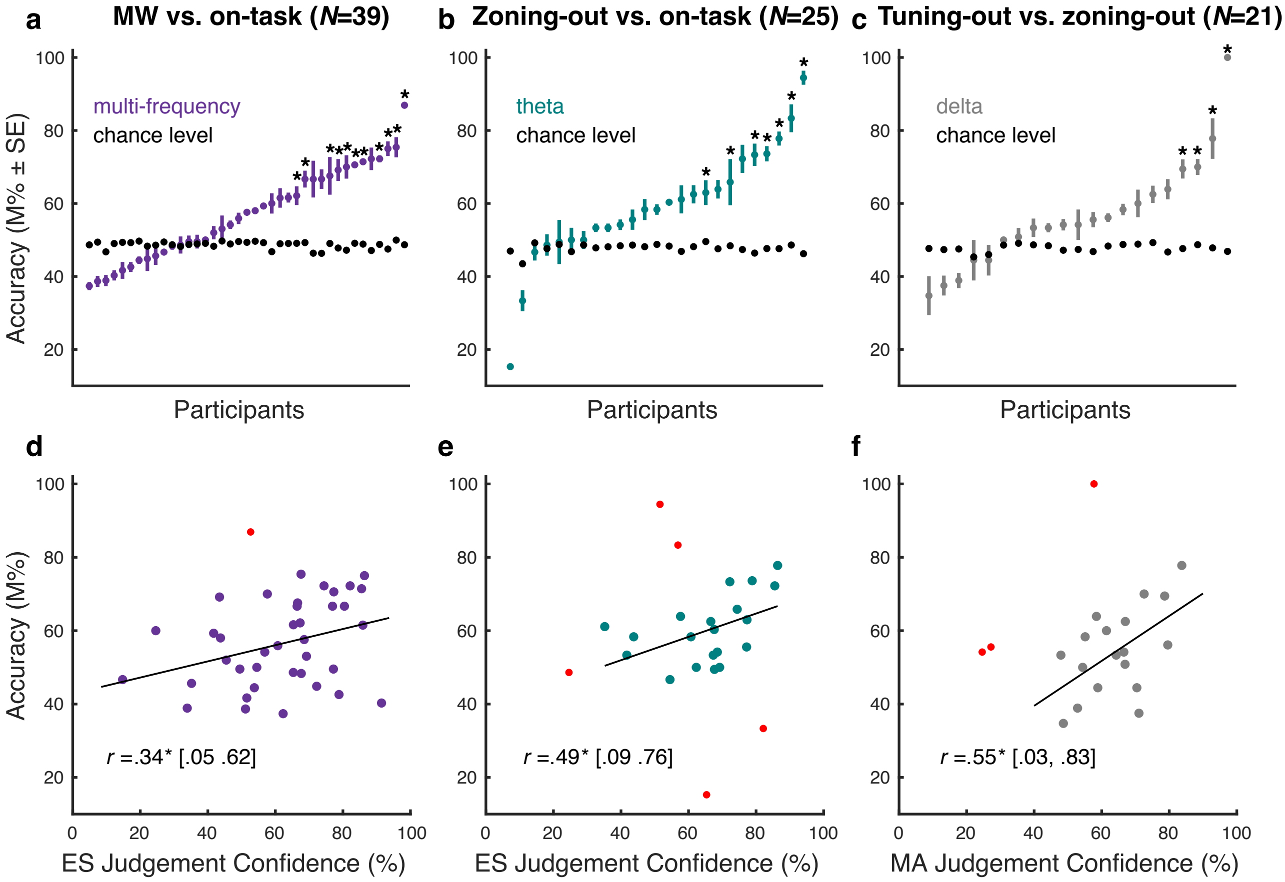


Figure 4. Time-averaged MVPC (Multivariate Pattern Classification analysis): (**a-c**) Decoding accuracy in individual participants relative to permutation distributions: chance level obtained by 500 random permutations after shuffling the class labels (upper panel). (**d-f**) Scatterplots of participant-level mean decoding accuracies and mean judgment confidence ratings. Red markers denote bivariate outliers; square brackets denote Bootstrap 95% CIs. MW = Mind Wandering.

* p<.05

2.4.3. Association between decoding accuracy and introspection confidence

Our final set of analyses were motivated by the hypothesis that introspection confidence reflects conscious accessibility to experiential states. Toward this end, we evaluated the predictions that individual decoding accuracies in select frequency bands would positively correlate with confidence in experiential state judgments. We computed correlations between decoding accuracies (multi-frequency model) and ES judgement confidence for on-task vs. mind wandering and tuning-out states, and decoding accuracies (theta and delta models) with MW and MA judgement confidence for zoning-out vs. on-task and tuning-out, respectively.

In support of our overarching prediction, decoding accuracies for MW vs. on-task (multi-frequency model) positively correlated with ES judgement confidence ratings, p=.036 (**Figure 4d**, [p=.08, with outliers]). By contrast, decoding accuracies for tuning-out vs. on-task (multi-frequency model) did not significantly correlate with ES judgement confidence, r=-.003, p=.99 [-0.36, 0.47]. Decoding accuracies for zoning-out vs. on-task (theta model) correlated positively with MW judgment confidence, p=.029 (**Figure 4e,** [p=.71, with outliers]), but was non-significant for the delta model, r=-.26, p=.22 [-.50, .06]. Finally, confidence in meta-awareness judgments positively correlated with decoding accuracies (delta model) for zoning-out vs. tuning-out, p=.018 (**Figure 4f**, [p=.32, with outliers]), but did not significantly correlate with decoding accuracies for the theta model, r=.27, p=.27 [-0.04, 0.57]. These results collectively suggest that experiential states rated with higher confidence are also more dissimilar at the neurophysiological level.

3 Discussion

Using EEG and an ecological listening task, this study investigated the neural oscillatory dynamics of mind wandering and meta-awareness states. Mind wandering was reliably characterized by greater alpha power than on-task states with more prominent effects in this band and both delta and theta bands for unaware mind wandering. Consistent with the notion that mind wandering is more pronounced when one lacks meta-awareness (Christoff et al. 2009), moment-to-moment variations between unaware mind wandering and on-task states were the most reliably decoded via multivariate pattern classification. Critically, we found that decoding accuracy in the classification of different experiential states predicted confidence in the corresponding state judgments. This is consistent with the proposal that confidence indexes metacognitive access to experiential states.

Mind wandering as external inattention and internal focus

As in previous research demonstrating the impact of mind wandering on comprehension (e.g. Boudewyn and Carter 2018), self-reported mind wandering was associated with poorer recall in the listening task. One of the principal results of this study was that power in the alpha frequency band was greater during mind wandering than on-task states, replicating previous work (Macdonald et al. 2011; Baldwin et al. 2017; Boudewyn and Carter 2018; Compton et al. 2019). This effect was specific to the alpha band and generalized across meta-awareness states suggesting that elevated alpha power is a frequency-specific but generalized neurophysiological characteristic of mind wandering. Our findings are in line with studies showing elevated alpha power at posterior sites when attention is focused internally (Cooper et al. 2003) as well as multiple lines of evidence suggesting that mind wandering is associated with decay of perceptual processing (Smallwood et al. 2008; Barron et al. 2011). These results suggest that greater alpha power reflects detachment from the external world and a shift towards internal processing (Smallwood and Schooler 2015).

Our findings suggest that unaware mind wandering is more divergent from on-task states than episodes of mind wandering with awareness [13]. Although both were characterized by higher alpha power compared to on-task states (Boudewyn and Carter, 2018), unaware mind wandering was additionally associated with greater delta and theta power, particularly in right frontal and parieto-central sites, respectively. Aware and unaware mind wandering had suggestively distinct oscillatory features thereby implying that state meta-awareness represents a dimension of attention that is orthogonal to the direction of attention. These findings may help to explain previous reports of elevated delta and theta during mind wandering in self-caught paradigms [32,33], which in our view almost exclusively index unaware mind wandering.

The observation of elevated theta and delta power during unaware mind wandering aligns with previous research on the cognitive correlates of oscillatory activity in these bands. Theta oscillations are suggested to be involved in cognitive control [61], working memory [34,36] and conflict detection [62], processes that could be implicated in appropriate selection between multiple simultaneous thoughts related to the processing of current concerns during unaware mind wandering. This aligns with hypothesised parallels between mind wandering and meditative states related to moment-to-moment navigation through mental objects [63]. In particular, our finding of higher theta power during unaware than aware mind wandering states is potentially congruent with higher frontal midline and temporo-parietal theta during meditative states characterised by deeper absorption [64] and thus suggest potential links between absorption and zone-outs. Delta frequency contributions have been revealed during increased focus on internal processing and pertinent inhibition of interference [35]. Delta and theta activity could thus reflect the involvement of memory in self-related processing during self-generated thoughts in the context of unaware mind wandering episodes. The complementary role of delta activity might be to preserve internal processing during unaware mind wandering by inhibiting external interference [35,65]. Accordingly, our findings may align with the proposal that mind wandering recruits processes to ensure that one's internal train of thought is maintained [66,67].

Decoding of experiential states

Although our data were not event-related, our time-varying multivariate classifier was able to trace mind wandering in several temporal segments within the 10 second time window preceding experiential state probes. Moreover, MVPC allowed us to decode experiential states from oscillatory activity at both participant- and group-levels, highlighting the utility of spectral measures coupled with machine learning in decoding mind wandering (Groot et al. 2021; Jin et al. 2019) and state meta-awareness. The time-varying and time-averaged analyses did not reveal substantially different results. In both analyses, all models decoded mind wandering from on-task states, including in approximately one-quarter of participants (except for the time-varying theta model) and with comparable classification accuracies. Similarly, aware mind wandering was decoded from on-task states in both analyses, with models using power in the delta band for classification showing the weakest classification performance. Time-varying analysis yielded significant decoding between unaware mind wandering and on-task states in multiple time windows, especially for delta and theta models, which decoded the two experiential states across almost the entire epoch. Comparably, the time-averaged MVPC achieved significance in more participants than any other classification, and also achieved the highest decoding accuracy (group-level). Finally, the time-averaged multi-frequency MVPC decoded aware from unaware mind wandering states in 29% of participants and the alpha model achieved the highest single-participant decoding (38%) for this classification.

Taken together, these results suggest that information pertaining to meta-awareness might be distributed across different frequency bands including slow oscillations – as shown in the higher prevalence of significant decoding in the sample. However, our analyses did not reveal any robust evidence for frequency specificity and thus it seems that there is not a specific oscillatory pattern that contains more information about mind wandering and state meta-awareness. We further corroborate that unaware mind wandering is more dissimilar at the neural level to external attention (on-task states) (Christoff et al. 2009). Combined with our MVPC results decoding aware from unaware mind wandering, we confirm that state meta-awareness should be considered as an important dimension of mind wandering in future studies, with evidence for disparate neural substrates for aware and unaware mind wandering. Collectively, MVPC trained on EEG-extracted features reliably decoded different experiential states both at participant and group levels. Discrepancies between group-level and participant-level classification is in line with research showing the utility of using individualized markers in decoding mind wandering [68].

Introspection confidence

Previous research investigating confidence suggests that self-reports of mind wandering characterised by higher confidence constitute more accurate evaluations of one's experiential states

(Seli et al. 2015). Our results build upon this, showing that confidence in experiential state judgments tended to map onto state meta-awareness: participants reported the greatest certainty for on-task episodes, were less confident in aware mind wandering, and the least for mind wandering episodes without awareness, although confidence ratings between the latter two were not significantly different. Future work could investigate how confidence relates to other prominent dimensions of mind wandering, such as intentionality [69,70]. Participant-level decoding of experiential states allowed us to evaluate the prediction that introspection confidence would be positively associated with decoding accuracy. Indeed, confidence in different experiential state judgments reliably correlated with individual differences in MVPC decoding accuracies. In particular, confidence correlated with cross-frequency decoding accuracy in classifying mind wandering from on-task states. In addition, consistent with findings implicating theta activity in metacognition [71], we also found that confidence in mind wandering judgments was associated with decoding of unaware mind wandering from on-task states accuracy in the theta model. Finally, confidence in meta-awareness judgements was positively associated with delta model accuracy in the decoding of aware versus unaware mind wandering. Collectively, these results demonstrate that confidence in one's experiential states is positively related to the multivariate decodability of these states with implications for the neural bases of experiential state confidence.

Our results align with previous findings [47] suggesting that high confidence levels reflect a more accurate assessment of one's experiential state reflected in stronger coupling between mind wandering reports and well-established impacts on behaviour. We extend this notion and provide evidence that higher confidence may reflect greater dissociation of experiential states at the neurophysiological level. In MVPC, higher decoding accuracy denotes better discriminability or separation between EEG patterns associated with each experiential state class. Although this discriminability refers only to pattern analysis, it is possible that more dissociable or distinct neural patterns are metacognitively represented and therefore reflected in individuals' confidence ratings. One interpretation of these findings thus is that states accompanied by high confidence are quantitatively more intense or salient and thus characterized by superior conscious access that is grounded in or related to underlying neurophysiological discriminability. Alternatively, this relationship might be attributable to participants with low confidence displaying weak experiential state discriminability, i.e., they have relatively poor metacognitive access in their experiential states resulting in lower multivariate decoding. However, whether and to what extent confidence judgements about internal states reflect a readout of discriminability between neural patterns remains an unresolved issue. Recent work demonstrated that confidence judgements and behavioural accuracy are dissociated during decision making and these dissociations can be explained by differences in neural computations [72]. One limitation in this study is that due to the small number of trials per certain classes in certain participants, our analyses were limited to mean confidence ratings. Future research on mind wandering could utilize confidence ratings on a trial-by-trial basis to provide a more precise estimate of the relationship between decoding accuracy and classification accuracy (Weaver et al. 2019).

Conclusions

Our findings expand upon research linking elevated alpha power with mind wandering episodes and reveal distinct electrophysiological characteristics of state meta-awareness. Unaware mind wandering was consistently more dissimilar from on-task states than aware mind wandering, as evidenced by superior decoding and greater neurophysiological differences. These results highlight a clear distinction between unaware and aware mind wandering states and confirm the utility of introspective methods in the study of transient fluctuations in conscious experience. The observed effects demonstrate the potential of using EEG machine learning classifiers to capture mind wandering and state meta-awareness during an ecological task without performance indicators. We found that confidence in experiential state reports correlated with the decoding of the respective states, suggesting that introspection confidence scales with neurophysiological dissimilarity. These effects suggest that introspection confidence taps into variability in metacognitive access to, and differential phenomenological characteristics, of experiential states.

4 Materials and Methods

4.1 Participants

Forty-six right-handed participants (28 females, age range: 18-43, M_{Age}=25.9, *SD*=5.7; years of education [post-secondary school]: M_{Yoe}=4.3, *SD*=2.2) with normal or corrected-to-normal vision provided written informed consent to volunteer in the study and were compensated £10 per hour. A sample size of 40 allowed us to detect paired-samples effects of $d \geq .45$ (α=.05, 1-β=.80, two-tailed). We recruited 46 participants due to potential attrition and loss of participants because of insufficient numbers of trials for the different state responses. All participants self-reported proficiency in English (1=no proficiency, to 10=native speaker; *M*=9.2, *SD*=1.04). Seven participants were excluded due to technical issues during EEG data recording (*n*=1), or insufficient number of response types in the task (*n*=6; see section 4.5), resulting in a final sample of 39 participants (25 females, age range: 18-43, M_{Age}=25.6, *SD*=5.8, English language skills: [*M*=9.2%, *SD*=1.0]). The study was approved by the Research Ethics Committee of the Department of Psychology at Goldsmiths, University of London.

4.2 Materials

Audiobook listening task. This task consisted of participants listening to an audio version of Bill Bryson's *A Short History of Nearly Everything* (2004), a general science book that has previously been used in mind wandering research (e.g. Smallwood, Nind, et al. 2009). Participants focused on a central white fixation cross on a grey background at a distance of approximately 90cm and listened (through speakers) to the audiobook in three 20min blocks (corresponding to chapters 7, 24, and 30 in counterbalanced order). During the task, participants were prompted via on-screen thought probes at pseudorandom intervals (30, 40, or 50s) to report on their experiential state: "Just before the probe, were you mind wandering?" (ES Judgement, response options: yes, no). If a participant responded in the affirmative, they were next prompted regarding state meta-awareness: "Just before the probe, were you zoning-out or tuning out?" (MA Judgement, response options: tuning-out, zoning-out). Participants responded to both probes using a continuous visual analogue scale in which they made their binary judgement combined with an estimate of confidence in their response (ranging from completely not confident to completely confident). Response options for both probes alternated sides randomly to control for response biases.

Audiobook listening assessment. A sequence of 20 true/false questions (corresponding to the content of the preceding block) were administered to participants after each block. Question order followed the presentation order of the information in the audiobook with each question corresponding to approximately 1 minute of content.

4.3 Procedure

After EEG preparation and general instructions, participants completed a battery of psychometric measures (to be reported elsewhere). Participants sat in a dimly lit room and first underwent a 5-min eyes-open resting state condition in which they focused on a central white fixation cross ($1cm^2$) whilst their EEG was recorded and subsequently completed a self-report resting state measure (to be reported elsewhere).

Prior to completing the task, mind wandering was defined to participants as any thoughts that are not related to the material being presented [13,74], and are usually internally focused. Participants were provided with examples of mind wandering, such as thoughts about past events, friends or significant others or concerns about an upcoming exam [57]. Tuning out was defined as a state in which one mind wanders and is aware whilst they are doing so whereas zoning out was defined as a state in which one mind wanders and is unaware that they are doing so until they "catch" themselves.

The experimenter introduced the audiobook listening task and defined the response options in the mind wandering task and ensured that participants understood these options. Participants completed a 3min training block followed by three 20min experimental blocks (30 probes per block), resulting in 90 probes in total. After each block, participants completed the audiobook listening assessment. Blocks took approximately 25 mins to complete with the entire experiment lasting approximately 2.5 hours. The experiment was programmed and implemented in MATLAB® (2018a, The MathWorks, Inc., MA), using the Psychophysics Toolbox extensions [75]. All data are available upon request.

4.4 Behavioural analyses

Participants' data were segregated at the probe-level according to self-report in two ways for separate analyses: dichotomously (on-task vs. mind wandering) and trichotomously (on-task vs. tuning-out vs. zoning-out). Frequency (%) of each state report and performance on the audiobook listening assessment (accuracy [%]) were additionally computed at the block-level.

4.5 Electrophysiological data acquisition and analyses

EEG signals were recorded using a 64-Ag-AgCl electrode Biosemi ActiveTwo system. Electrodes were placed according to the International 10-20 system. Two electrodes placed on the participants' earlobes were used as reference. Additional electrodes recorded right side vertical (VEOG) and bilateral horizontal (HEOG) electro-oculogram signals to be used for artefact detection and rejection. The recording was sampled at 512Hz for all participants.

Data pre-processing was implemented using the EEGlab toolbox in MATLAB [76]. The average of the two earlobe electrodes was used as reference; the data were subsequently filtered with a high-pass filter at 0.5Hz and a notch-filter between 48-52Hz. We used the pop_eegfiltnew function in EEGlab, which applies a finite impulse response (FIR) filter to the data using an automatic filter order (3380 and 846, respectively). Bad electrodes (range: 0-2 across participants) detected during the recording, or via visual inspection of the raw data, were removed. Next, independent component analysis (ICA) was performed on the continuous data to detect eye-movement artefacts. IC scalp maps, spectra and raw activity were visually inspected to reject further artefacts such as eye movements, further channel noise, and prominent muscle movements. Next, data from removed electrodes were replaced using spherical interpolation and all data were re-referenced using the average of the 64 channels.

For each participant, continuous data were next segmented into 14s epochs: -12 to 2s relative to probe onset. The time window of interest extended from -10 to 0s, but an additional 2s was included on each side to avoid edge artefacts in subsequent analyses; these data were omitted after time-frequency transformation. Further epoch exclusion was conducted via manual rejection based on visual inspection. Data were subsequently segregated into two conditions corresponding to the experiential states reported by the participants: on-task vs. mind wandering. Mind wandering states were further partitioned into two meta-awareness states: tuning-out vs. zoning-out. A time-frequency transformation of the data was implemented by applying a Hanning window at 50ms steps to each 14s-long epoch and corresponding baseline segment for frequencies of 1 to 45Hz. Window length varied along the frequency dimension, with 7000ms at the lowest frequency (1Hz) decreasing linearly (time window=7/frequency) at each frequency bin, and 150ms for 45Hz. Trial-wise spectral power was averaged and then normalised by division of a baseline level (gain model; Grandchamp and Delorme 2011). We used as baseline a 700ms epoch during the inter-stimulus interval after the probe response phase and start of the next trial.

Participants varied in their mind wandering and meta-awareness reports, resulting in different sample sizes and respective numbers of trials per state for each comparison. State-specific data that included fewer than 10% of probes (9 trials) were excluded from any analyses involving the respective state. Four main contrasts were implemented with variable trials (M±SD) and sample sizes [N]: (i) on-task (39.7±15.0) vs. mind wandering (27.4±12.0) [N=39]; (ii) on-task (36.0±13.0) vs. tuning-out (20.5±9.8) [N=27]; (iii) on-task (35.5±13.7) vs. zoning-out (14.1±5.4) [N=25]; and (iv) tuning-out (22.0±13.4) vs. zoning-out (13.3±4.0) [N=21].

4.6 Multivariate pattern classification analysis (MVPC)

Complementing the univariate analysis of spectral power differences, we implemented a two-class MVPC for the four aforementioned two-state contrasts. Here, we hypothesized that information about the different experiential states would be shared across different frequency bands and electrodes. Insofar as this information could unfold over time, we used time-varying MVPC to investigate whether subjective reports about experiential states could be decoded from trial-wise EEG patterns of oscillatory activity across different frequency bands, and separately for different time points.

Our analyses were based on trial-wise measures of spectral power for delta, theta and alpha frequency bands and for each of the 64 EEG channels. These measures were averaged in bins of 500ms within 10s epochs, resulting in 21 time bins. A support vector machine (SVM library for MATLAB; Lotte et al. 2007; Chang and Lin 2011) was trained to distinguish between classes (states) at each time bin. Each of the implemented two-class MVPC analyses were balanced by matching the number of trials in each class using semi-random trial selection. We used 3-fold cross-validation to obtain estimates of decoding accuracy at the participant-level. In order to examine whether single-frequency classification was superior or comparable to cross-frequency classification (multi-frequency model), we performed control MVPC analyses separately for each frequency band (reduced feature space).

An important consideration is that, although mind wandering processes could unfold over time, they are inherently not time-locked, and indeed our experimental design was not event-related. To account for the case that there would not be a reliable temporal representation of mind wandering states across trials, we further complemented the time-varying MVPC with a time-averaged MVPC approach. We collapsed the spectral power measures across time to assess whether we could still decode classes from the same spectral and spatial features. An additional advantage of this second approach is that using time-averaged measures of spectral power in MVPC reduces the number of multiple comparisons. Spectral power at each time point was averaged across the whole epoch (−10 to 0s before probe onset). Similar to the time-varying analyses, the time-averaged MVPC was performed for the multi-frequency model and separately for each frequency band of interest (single-frequency models). Finally, we conducted correlation analyses between decoding accuracies derived from our time-averaged models and participants' confidence in their respective reports. Correlation analyses were performed exclusively in a subset of the models, based on our EEG findings and previous research (see section 4.7.3).

4.7 Statistical analyses

4.7.1 Behavioural data

Confidence ratings between states were compared with paired-samples *t*-tests (two-tailed) and assessment performance was compared to 50% using a one-sample *t*-test. Associations between task-level mind wandering frequency and assessment accuracy were assessed by correlational analyses following automatic bivariate outlier (boxplot method) removal using the Robust Correlation toolbox in MATLAB [80] in the computation of skipped correlations [81–83]. We report Spearman's r_s for data that violated parametric test assumptions.

4.7.2 Spectral analysis

Significant differences in the spectral power between different states were assessed by means of cluster-based paired permutation tests across participants [84]. Non-parametric cluster permutation tests were used separately for pre-specified oscillatory frequency bins (delta [2-3Hz], theta [4-7Hz], alpha [8-13Hz]). These analyses were undertaken in two phases. First, we calculated the observed test statistic for the respective contrast by: (i) conducting paired-samples *t*-tests comparing the two states at each data sample (frequency x channel x time); (ii) samples whose *t*-values were below threshold ($\alpha<.05$) were selected and clustered in sets based on feature adjacency (spectral, spatial, and temporal); and (iii) *t*-values were summed to compute cluster-level statistics whose maximum served as the test statistic to evaluate state differences. The second phase entailed the same steps but this time the test statistic was computed for 500 permutations of randomly partitioned data in two subsets (Monte Carlo permutation test). These test statistics were compared to the observed test statistic [84]. The cluster-level significance value was set at two-tailed $\alpha<.025$ with a minimum of 2 neighbouring channels constituting a cluster. This method controls for multiple comparisons by controlling the family wise error rate at $\alpha=.05$ We interpret multiple effects in the range of $.025<p<.030$ as reflecting trends. The analyses were conducted with the Fieldtrip toolbox in MATLAB [84]. Effect sizes were estimated using Hedges's *g* and bootstrap 95% confidence intervals (CI, bias-corrected and accelerated method, 10000 samples [Efron 1987]), on power averages across frequency, time window, and electrode sites identified by cluster analyses, using the Measures of Effect Size Toolbox in MATLAB (Hentschke 2021).

4.7.3 MVPC analyses

For the time-varying MVPC analyses, we initially performed participant-level classification and later estimated group-level decoding accuracy averaged across participants. At the participant level, the null distribution for accuracy was computed by performing the analysis 500 times after randomly shuffling the class labels in the data. *P*-values were computed at each time bin as the proportions (%) of permutation accuracies that are greater than or equal to the observed decoding accuracy (mean accuracy of 3-fold cross-validation), yielding one *p*-value per time bin. We report the proportion of participants for which we observed significant decoding in at least one time bin. Next, statistical assessment with a permutation test was performed at the group-level; we compared group mean accuracies from cross-validation and the estimated null distribution (Monte Carlo permutation test, 5000 permutations) to identify the time bins showing statistically significant decoding accuracy across participants. Both at the participant and group-levels, we corrected for multiple comparisons by controlling the false discovery rate (FDR) at .05 by using an adaptive two-stage linear step-up procedure [86]. At the group-level, the corrected threshold *p*-value obtained from this procedure, *p*th, is given when multiple comparisons were performed.

Time-averaged MVPC analyses were performed in the same manner but were limited to a single averaged time bin. At the participant level, the null distribution for accuracy was computed by performing the analysis 500 times after randomly shuffling the class labels in the data. *P*-values were computed as the proportions (%) of permutation accuracies that are greater than or equal to the observed decoding accuracy (mean accuracy of 3-fold cross-validation), yielding one *p*-value per participant. We report the percentage of participants for which we observed significant ($p<.05$) decoding for each model (multifrequency, single-frequency), in addition to the mean observed accuracies at the group-level. Different models were compared using the Monte Carlo approach described above (paired permutation test). Finally, we assessed associations between mean participant-level decoding accuracies for each comparison and participants' corresponding mean confidence ratings. To minimize the family-wise error rate, we selected a small number of models for this analysis based on the previous literature and our EEG data. We examined the relationship between decoding accuracies in the multi-frequency model and ES judgement confidence for on-task vs. mind wandering and tuning-out states, and theta and delta models' decoding accuracies with ES and MA judgement confidence for zoning-out vs. on-task and tuning-out respectively. As described above, bivariate outliers were removed in the computation of skipped Pearson correlations [81–83].

Data Availability: The data that support the findings of this study are available from the corresponding author upon request.

Code Availability: Custom code used for MVPC analyses is available from the corresponding author upon request.

5 References

1. Antrobus, J. S., Singer, J. L., Goldstein, S. & Fortgang, M. Section of Psychology: Mind Wandering and Cognitive Structure. *Trans. N. Y. Acad. Sci.* **32**, 242–252 (1970).

2. Smallwood, J. & Schooler, J. W. The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).

3. Killingsworth, M. A. & Gilbert, D. T. A wandering mind is an unhappy mind. *Science (80-.).* **330**, 932 (2010).

4. 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).

5. Kam, J. W. Y. *et al.* Slow Fluctuations in Attentional Control of Sensory Cortex. *J. Cogn. Neurosci.* **23**, 460–470 (2011).

6. Baird, B., Smallwood, J., Lutz, A. & Schooler, J. W. The Decoupled Mind: Mind-wandering Disrupts Cortical Phase-locking to Perceptual Events. *J. Cogn. Neurosci.* **26**, 2596–2607

(2014).

7. Geden, M., Staicu, A.-M. & Feng, J. The impacts of perceptual load and driving duration on mind wandering in driving. *Transp. Res. Part F Traffic Psychol. Behav.* (2017) doi:10.1016/j.trf.2017.07.004.

8. Feng, S., D'Mello, S. & Graesser, A. C. Mind wandering while reading easy and difficult texts. *Psychon. Bull. Rev.* **20**, 586–592 (2013).

9. Hollis, R. B. & Was, C. A. Mind wandering, control failures, and social media distractions in online learning. *Learn. Instr.* **42**, 104–112 (2016).

10. Schooler, J. W., Reichle, E. D. & Halpern, D. V. Zoning Out while Reading: Evidence for Dissociations between Experience and Metaconsciousness. in *Thinking and seeing: Visual metacognition in adults and children* 203–226 (MIT Press, 2004).

11. Raichle, M. E. *et al.* A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 676–82 (2001).

12. Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9673–8 (2005).

13. 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. U. S. A.* **106**, 8719–24 (2009).

14. Fingelkurts, A. A. & Fingelkurts, A. A. Persistent operational synchrony within brain default-mode network and self-processing operations in healthy subjects. *Brain Cogn.* **75**, 79–90 (2011).

15. Andrews-Hanna, J. R. The Brain's Default Network and Its Adaptive Role in Internal Mentation. *Neurosci.* **18**, 251–270 (2012).

16. Andrews-Hanna, J. R., Smallwood, J. & Spreng, R. N. The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* **1316**, 29–52 (2014).

17. Spreng, R. N. & Grady, C. L. Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *J. Cogn. Neurosci.* **22**, 1112–1123 (2010).

18. 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**, 527–33 (2007).

19. 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).

20. Martinon, L. M., Smallwood, J., McGann, D., Hamilton, C. & Riby, L. M. The disentanglement of the neural and experiential complexity of self-generated thoughts: A users guide to combining experience sampling with neuroimaging data. *Neuroimage* **192**, 15–25 (2019).

21. Boudewyn, M. A. & Carter, C. S. I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia* **117**, 148–155 (2018).

22. Compton, R. J., Gearinger, D. & Wild, H. The wandering mind oscillates: EEG alpha power is enhanced during moments of mind-wandering. *Cogn. Affect. Behav. Neurosci.* **19**, 1184–1191 (2019).

23. Groot, J. M. *et al.* Probing the neural signature of mind wandering with simultaneous fMRI-EEG and pupillometry. *Neuroimage* **224**, (2021).

24. Macdonald, J. S. P., Mathan, S. & Yeung, N. Trial-by-trial variations in subjective

attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Front. Psychol.* **2**, 82 (2011).

25. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **16**, 606–617 (2012).

26. Palva, S. & Palva, J. M. New vistas for α-frequency band oscillations. *Trends Neurosci.* **30**, 150–158 (2007).

27. Foxe, J. J. & Snyder, A. C. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* **2**, 1–13 (2011).

28. Von Stein, A. & Sarnthein, J. Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *Int. J. Psychophysiol.* **38**, 301–313 (2000).

29. Cooper, N. R., Croft, R. J., Dominey, S. J. J., Burgess, A. P. & Gruzelier, J. H. Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *Int. J. Psychophysiol.* **47**, 65–74 (2003).

30. Hanslmayr, S., Gross, J., Klimesch, W. & Shapiro, K. L. The role of alpha oscillations in temporal attention. *Brain Res. Rev.* **67**, 331–343 (2011).

31. Villena-González, M., López, V. & Rodríguez, E. Orienting attention to visual or verbal/auditory imagery differentially impairs the processing of visual stimuli. *Neuroimage* **132**, 71–78 (2016).

32. Braboszcz, C. & Delorme, A. Lost in thoughts: Neural markers of low alertness during mind wandering. *Neuroimage* **54**, 3040–3047 (2011).

33. van Son, D. *et al.* Frontal EEG theta/beta ratio during mind wandering episodes. *Biol. Psychol.* **140**, 19–27 (2019).

34. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a rKlimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews, 29(2-3), 169–195. doi:10.1016/S016. *Brain Res. Rev.* **29**, 169–195 (1999).

35. Harmony, T. The functional significance of delta oscillations in cognitive processing. *Front. Integr. Neurosci.* **7**, 1–10 (2013).

36. Mitchell, D. J., McNaughton, N., Flanagan, D. & Kirk, I. J. Frontal-midline theta from the perspective of hippocampal 'theta'. *Prog. Neurobiol.* **86**, 156–185 (2008).

37. Palva, J. M., Palva, S. & Kaila, K. Phase synchrony among neuronal oscillations in the human cortex. *J. Neurosci.* **25**, 3962–3972 (2005).

38. Sauseng, P. *et al.* EEG alpha synchronization and functional coupling during top-down processing in a working memory task. *Hum. Brain Mapp.* **26**, 148–155 (2005).

39. Smallwood, J. & Schooler, J. W. The restless mind. *Psychol. Bull.* **132**, 946–58 (2006).

40. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).

41. Morgan, M. J., Mason, A. J. S. & Solomon, J. A. Blindsight in normal subjects? [10]. *Nature* **385**, 401–402 (1997).

42. Kunimoto, C., Miller, J. & Pashler, H. Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* **10**, 294–340 (2001).

43. Gherman, S. & Philiastides, M. G. Human VMPFC encodes early signatures of confidence in perceptual decisions. *Elife* **7**, 1–28 (2018).

44. Boldt, A. & Yeung, N. Shared neural markers of decision confidence and error detection. *J. Neurosci.* **35**, 3478–3484 (2015).

45. Kubanek, J., Hill, N. J., Snyder, L. H. & Schalk, G. Cortical alpha activity predicts the confidence in an impending action. *Front. Neurosci.* **9**, 1–15 (2015).

46. Zakrzewski, A. C., Wisniewski, M. G., Iyer, N. & Simpson, B. D. Confidence tracks sensory- and decision-related ERP dynamics during auditory detection. *Brain Cogn.* **129**, 49–58 (2019).

47. Seli, P., Jonker, T. R., Cheyne, J. A., Cortes, K. & Smilek, D. Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 703–709 (2015).

48. Meier, M. E. Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? A replication and extension of Seli, Jonker, Cheyne, Cortes, and Smilek (2015). *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 1567–1585 (2018).

49. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (80-.).* **293**, 2425–2430 (2001).

50. Haynes, J. D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).

51. 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).

52. Kim, J. *et al.* Decoding accuracy in supplementary motor cortex correlates with perceptual sensitivity to tactile roughness. *PLoS One* **10**, 1–17 (2015).

53. Weaver, M. D., Fahrenfort, J. J., Belopolsky, A. & Van Gaal, S. Independent neural activity patterns for sensory-and confidence-based information maintenance during category-selective visual processing. *eNeuro* **6**, (2019).

54. Smallwood, J., Nind, L. & O'Connor, R. C. When is your head at? An exploration of the

factors associated with the temporal focus of the wandering mind. *Conscious. Cogn.* **18**, 118–125 (2009).

55. Baldwin, C. L. *et al.* Detecting and quantifying mind wandering during simulated driving. *Front. Hum. Neurosci.* **11**, 1–15 (2017).

56. Allefeld, C. & J-D Haynes. *Multi-voxel Pattern Analysis*. *Brain Mapping: An Encyclopedic Reference* vol. 1 (Elsevier Inc., 2015).

57. Varao Sousa, T. L., Carriere, J. S. A. & Smilek, D. The way we encounter reading material influences how frequently we mind wander. *Front. Psychol.* **4**, 892 (2013).

58. Broadway, J. M., Franklin, M. S. & Schooler, J. W. Early event-related brain potentials and hemispheric asymmetries reveal mind-wandering while reading and predict comprehension. *Biol. Psychol.* **107**, 31–43 (2015).

59. Barron, E., Riby, L. M., Greer, J. & Smallwood, J. Absorbed in Thought. *Psychol. Sci.* **22**, 596–601 (2011).

60. 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).

61. Cavanagh & Frank, M. J. Frontal Theta as a Mechanism for Affective and Effective Control. *Trends Cogn. Sci.* **18**, 414–421 (2014).

62. Cohen, M. X. A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci.* **37**, 480–490 (2014).

63. Vago, D. R. & Zeidan, F. The brain on silent: mind wandering, mindful awareness, and states of mental tranquility. *Ann. N. Y. Acad. Sci.* **1373**, 96–113 (2016).

64. DeLosAngeles, D. *et al.* Electroencephalographic correlates of states of concentrative

meditation. *Int. J. Psychophysiol.* **110**, 27–39 (2016).

65. Harmony, T. *et al.* EEG delta activity: An indicator of attention to internal processing during performance of mental tasks. *Int. J. Psychophysiol.* **24**, 161–171 (1996).

66. Smallwood, J., Fishman, D. J. & Schooler, J. W. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychon. Bull. Rev.* **14**, 230–236 (2007).

67. Smallwood, J., Brown, K., Baird, B. & Schooler, J. W. Cooperation between the default mode network and the frontal-parietal network in the production of an internal train of thought. *Brain Res.* **1428**, 60–70 (2012).

68. Dhindsa, K. *et al.* Individualized pattern recognition for detecting mind wandering from EEG during live lectures. *PLoS One* **14**, 1–30 (2019).

69. Seli, P., Risko, E. F., Smilek, D. & Schacter, D. L. Mind-Wandering With and Without Intention. *Trends Cogn. Sci.* **20**, 605–17 (2016).

70. 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).

71. Wokke, M. E., Cleeremans, A. & Ridderinkhof, K. R. Sure i'm sure: Prefrontal oscillations support metacognitive monitoring of decision making. *J. Neurosci.* **37**, 781–789 (2017).

72. Peters, M. A. K. *et al.* Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat. Hum. Behav.* **1**, 1–8 (2017).

73. Bryson, B. *A short history of nearly everything*. (Black Swan, 2004).

74. Lindquist, S. I. & McLean, J. P. Daydreaming and its correlates in an educational environment. *Learn. Individ. Differ.* **21**, 158–167 (2011).

75. Kleiner, M. *et al. Perception. Perception* vol. 36 ([Pion Ltd.], 2007).

76. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

77. Grandchamp, R. & Delorme, A. Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Front. Psychol.* **2**, 1–14 (2011).

78. Chang, C. C. & Lin, C. J. LIBSVM: A Library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).

79. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. & Arnaldi, B. A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* **4**, (2007).

80. Pernet, C. R., Wilcox, R. & Rousselet, G. A. Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Front. Psychol.* (2013) doi:10.3389/fpsyg.2012.00606.

81. Rousseeuw, P. J. & Van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223 (1999).

82. Rousseeuw, P. J. Least median of squares regression. *J. Am. Stat. Assoc.* **79**, 871–880 (1984).

83. Verboven, S. & Hubert, M. LIBRA: A MATLAB library for robust analysis. *Chemom. Intell. Lab. Syst.* **75**, 127–136 (2005).

84. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

85. Efron, B. Better bootstrap confidence intervals. *J. Am. Stat. Assoc.* **82**, 171–185 (1987).

86. Benjamini, Y., Krieger, A. M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507 (2006).

Author Contributions: N.P and D.B.T initiated the project, and all authors designed the project. N.P. conducted the experiment and analyses. M.H.R provided the code for the MVPC analyses and MVPC figures. N.P. drafted, and all authors reviewed the manuscript.

Competing interests: All authors declare that they have no competing interests.

Supplementary Materials

Introspection confidence predicts EEG decoding of self-generated thoughts and meta-awareness.

Naya Polychroni[1, a], Maria Herrojo Ruiz[1,2], & Devin B. Terhune [1]

[1] *Department of Psychology, Goldsmiths, University of London, UK*

[2] *Center for Cognition and Decision Making, National Research University Higher School of Economics, Russian Federation*

[a] Corresponding author:

Naya Polychroni

Department of Psychology

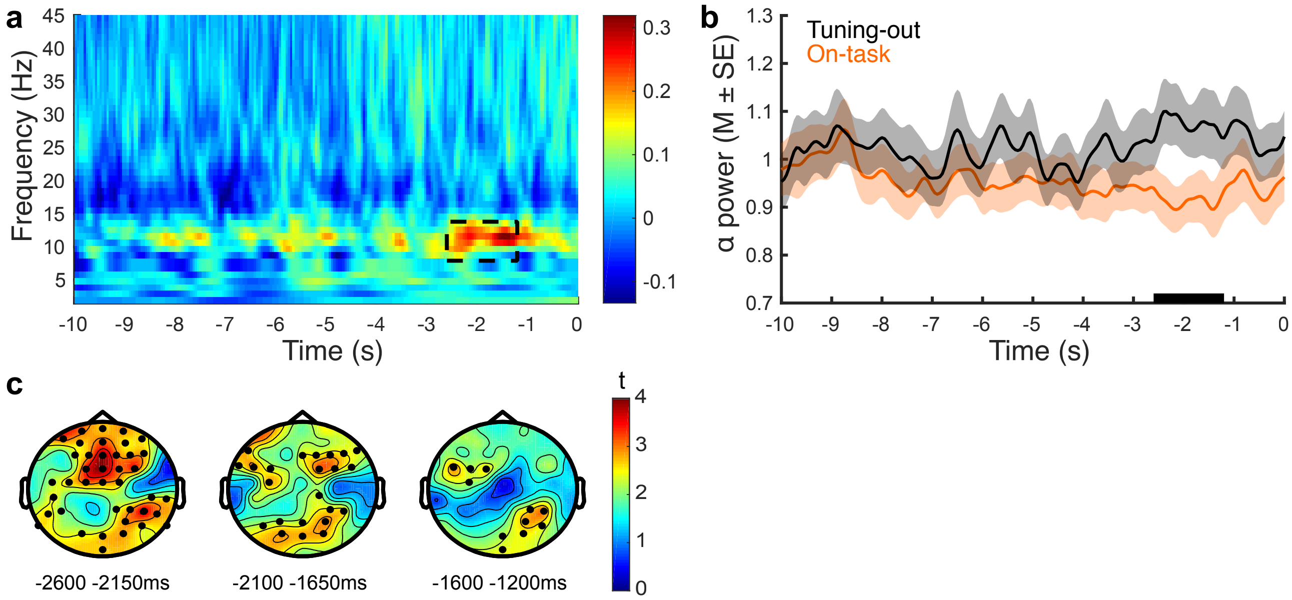
Goldsmiths, University of London

8 Lewisham Way

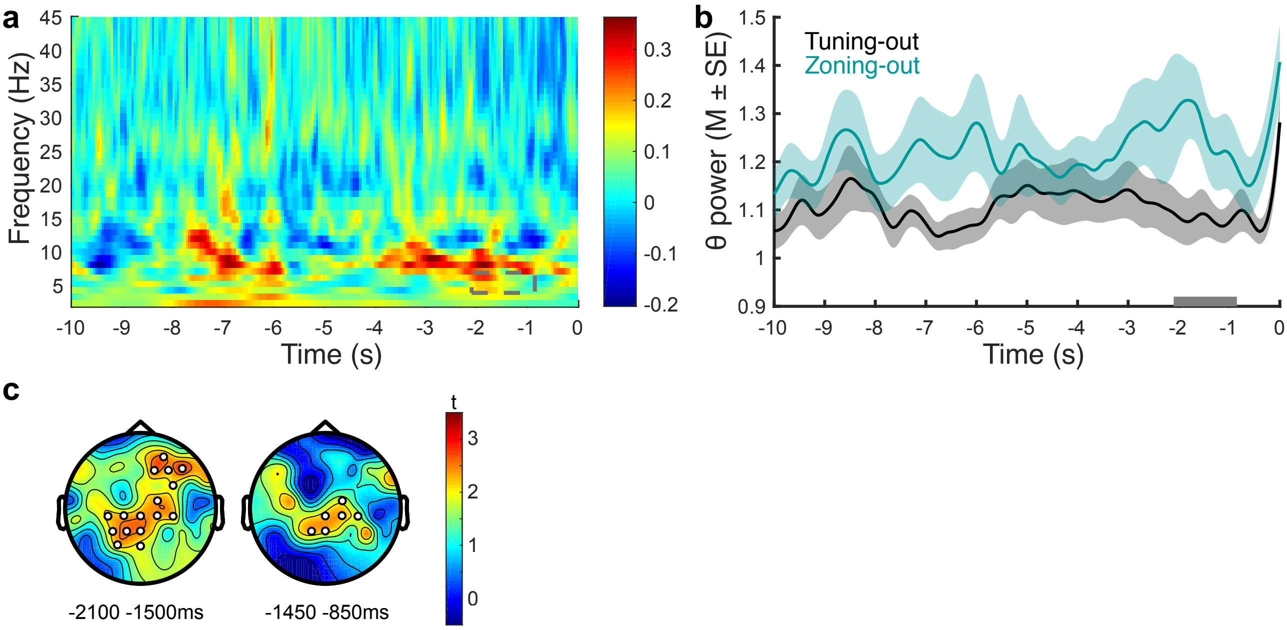
New Cross, London, UK SE14 6NW

ppoly002@gold.ac.uk

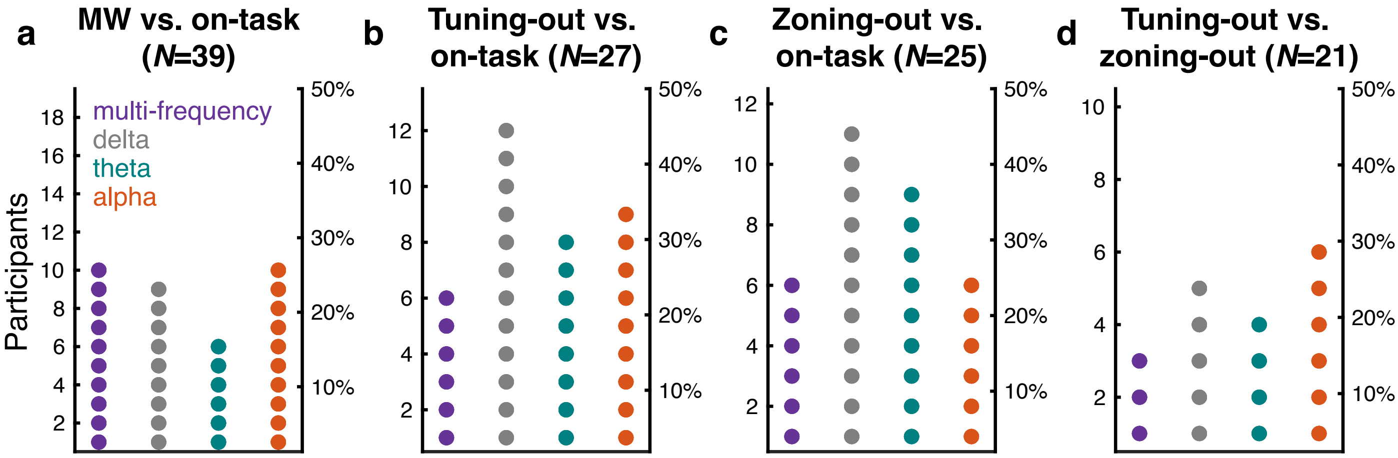
Supplementary Results



Supplementary Figure 1. Oscillatory differences between states (on-task – tuning-out, *N*=27) as a function of time relative to probe onset (0s). a) Time frequency decomposition averaged across all electrode sites. The broken black rectangle indicates the spectrotemporal cluster reflecting significant state differences (*p*<.025, two-sided cluster-based permutation test). b) Alpha (8-13Hz) spectral power averaged over the electrode sites of the cluster (significance denoted by a black bar on the x-axis). c) Topography of the cluster at different 400/450ms sub-windows (black markers denote electrodes that were present in at least 50% of samples in each time window).



Supplementary Figure 2. Oscillatory differences between states (zoning-out - tuning-out, *N*=21) as a function of time relative to probe onset (0s). a) Time frequency decomposition averaged across electrode sites. The broken grey rectangle denotes a spectrotemporal cluster reflecting the largest identified state differences (*p*=.026, two-sided cluster-based permutation test). b) Theta (4-7Hz) spectral power averaged over the electrode sites of the cluster (trend effect denoted by grey bar on the x-axis). c) Topography of the cluster at two 600ms sub-windows (white markers denote electrodes that were present at least 50% of samples in time window).



Supplementary Figure 3. Time-varying MVPC (Multivariate Pattern Classification analysis): Participants (count and percentage) with significant (*p<pth*) decoding in at least one time bin. MW = Mind Wandering.

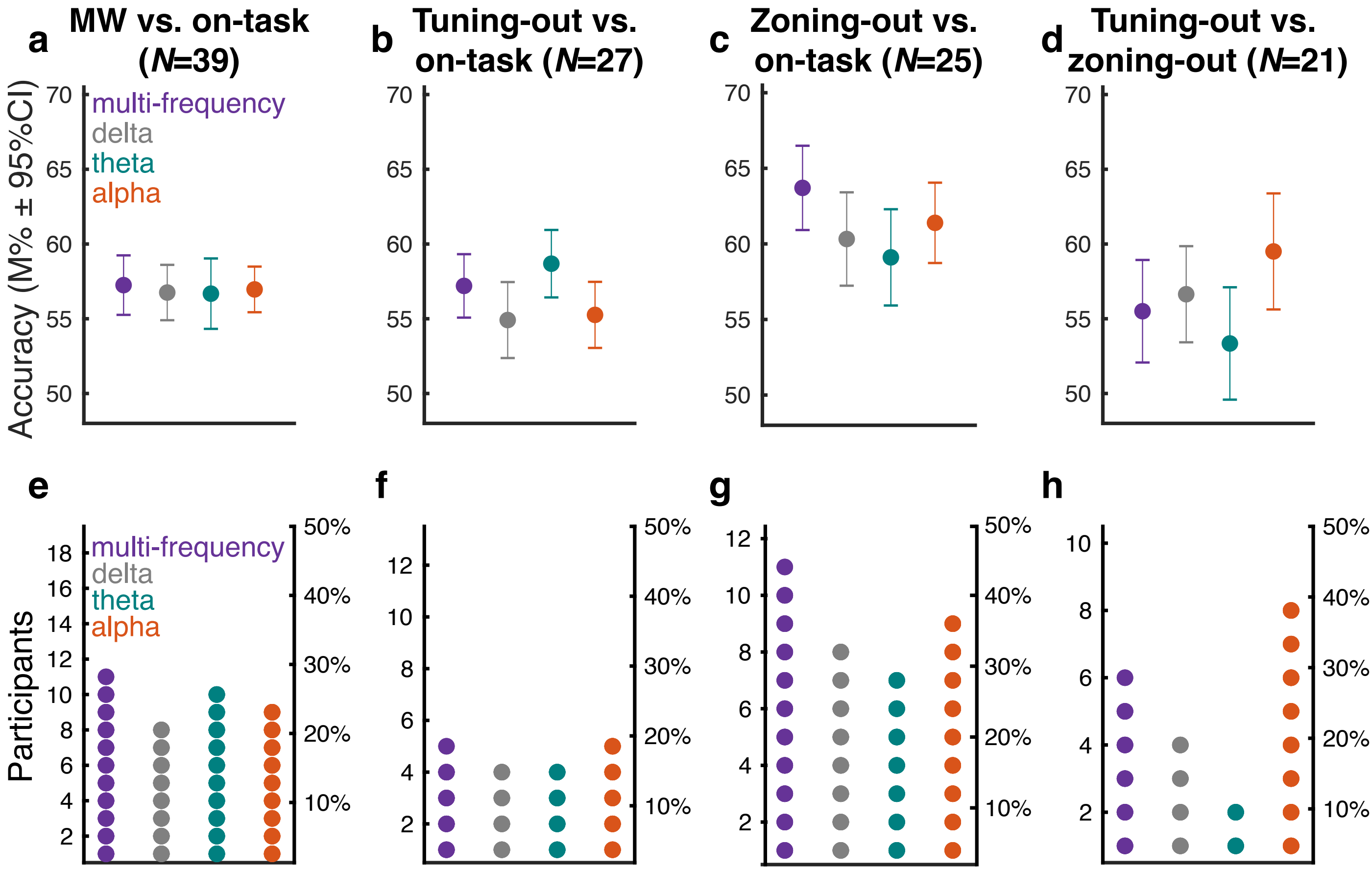
Supplementary Table 1

Time-varying MVPC (Multivariate Pattern Classification analysis) – Group-level analyses FDR (False Discovery Rate) threshold p-values

	Multi-frequency	Delta	Theta	Alpha
MW vs. on-task (N=39)	$p \leq pth$=.002	$p \leq pth$=.026	$p \leq pth$ =.018	$p \leq pth$ =.008
Tuning-out vs. on-task (N=27)	$p \leq pth$ =.016	ns, $p > pth$	$p \leq pth$ =.017	$p \leq pth$ =.005
Zoning-out vs. on-task (N=25)	$p \leq pth$=.003	$p \leq pth$ =.300	$p \leq pth$=.037	$p \leq pth$ <.001
Zoning-out vs. tuning-out (N=21)	$p \leq pth$<.001	$p \leq pth$=.016	$p \leq pth$=.021	$p \leq pth$ =.006

* These correspond to Figure 3 of the main text.

* MW = Mind Wandering.



Supplementary Figure 4. Time-averaged MVPC (Multivariate Pattern Classification analysis): **(a-d)** Group decoding accuracy across participants in multi-frequency and single-frequency models for four different state contrasts. (**d-f**) Participants (count and percentage) with significant (p<.05) decoding. MW = Mind wandering.