Context-dependent neural effects predict mind wandering minutes in advance

Chie Nakatani[1,*], Hannah Bernhard[1,2,3], & Cees van Leeuwen[1, 4]

Affiliations:
1: Brain and Cognition Research Unit, KU Leuven, Tiensestraat 102 - Box 3711, 3000 Leuven, Belgium

2: Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands (Current affiliation of H. Bernhard)

3: Centre for Integrative Neuroscience, Maastricht University, Maastricht, The Netherlands (Current affiliation of H. Bernhard)

4: Center for Cognitive Science, TU Kaiserslautern, Kaiserslautern, Germany

*Corresponding author: Chie Nakatani Email: chie.nakatani@kuleuven.be

Authors' Email: chie.nakatani@kuleuven.be, hannah.bernhard@maastrichtuniversity.nl,

cees.vanleeuwen@kuleuven.be

Key words: Spontaneous mental process, cognitive control, default mode network, dynamical model, delay, attention, working memory, alpha amplitude, evoked potentials

Funding: This work was supported by an Odysseus grant from The Flemish Organization for Science PEP-C3738-G.0003.12 to CvL

Running head: Predicting mind wandering

Abstract

Mind wandering occurs when spontaneous thought generation increases while cognitive control decreases. These are complex psychological events involving multiple brain systems, notably the default-mode (DMN) network and an antagonist system of cognitive control/frontoparietal attention networks, which are here addressed collectively as goal-directed networks (GDN). As opposed to the task-positivity of GDN, DMN exhibits largely task-negative effects. However, DMN could become task-positive when endogenous thoughts dominate. We hypothesized that a switch from task-negativity to positivity could be a factor in the transition between task-focused and mind wandering states. We estimated stochastic causality of the DMN- and GDN activity on mind wandering behavior, in a tone counting task designed to induce mind wandering in healthy adult participants. Electroencephalogram (EEG) was recorded together with subjective reports of mind wandering and other psychological and behavioral measures. Stochastic causality was found between measures of DMN activity (alpha-band power) and cognitive control (% correct of the task) prior to both task focused and mind wandering behavior. The estimated delay of the neural effect on the behavioral measure was 4-10 min. The effect was task-negative in task focused, but positive in mind wandering conditions. The results suggest that DMN slowly but dynamically influences cognitive control, through which the transition from task focused to mind wandering behavior is effectuated.

Introduction

The wakeful mind has a tendency to continuously generate thoughts, including ones that are irrelevant to the present situation, and may trigger episodes of *mind wandering* (Antrobus, 1968; Giambra, 1995; Klinger & Cox, 1987; Smallwood & Schooler, 2006). This propensity is countered by cognitive control functions that support intentional and goal-directed behavior (Aben, Verguts, & Van den Bussche, 2017; Cohen, Dunbar, & McClelland, 1990; Funes, Lupiáñez, & Humphreys, 2010; Otto, Skatova, Madlon-Kay, & Daw, 2015; Ridderinkhof, 2002). Whereas external circumstances may prompt instantaneous shifts between mind wandering and focused states, those states may also alternate spontaneously, as both thought generation and cognitive control strength vary over time. When many thoughts arise while cognitive control decreases, a shift to mind wandering becomes likely (Nakatani, Ganschow, & van Leeuwen, 2019).

These fluctuations could be understood as a function of endogenous, oscillatory brain rhythms reflecting the antagonism of two neural systems. One consists of the set of networks that typically show higher activity when performing demanding attention/working memory tasks than in resting conditions. This system encompasses the dorsal attention network, including the dorsolateral prefrontal cortex and the intraparietal sulcus (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Hopfinger, Buonocore, & Mangun, 2000), the salience network, including the insula and anterior cingulate cortex (Menon & Uddin, 2010; Seeley et al., 2007), and the sensory-motor systems. Increased activity in different subsystems of the GDN characterizes various forms of goal-directed behavior (Cole & Schneider, 2007). As a collective, we will call them the Goal-Directed Networks (GDN).

The opposing system is known as the Default Mode Network (DMN). This system encompasses widespread brain regions including prefrontal cortex, posterior cingulate cortex, as well as precuneus and memory-related regions such as the parahippocampal and middle temporal gyri (Raichle, 2015; Raichle et al., 2001). DMN regions generally show higher activity in resting conditions than during task performance. Such task-negativity was observed for a variety of neural measures, including the ratio of oxygen consumption to blood flow in PET (Raichle et al., 2001), BOLD in fMRI (Greicius, Krasnow, Reiss, & Menon, 2003), intracranial (stereo) EEG (Jerbi et al., 2010), source-estimated scalp EEG (Laufs et al., 2003;

Mantini, Perrucci, Del Gratta, Romani, & Corbetta, 2007) and MEG (Brookes et al., 2011). DMN regions tend to show higher activation during mind wandering than in task-focused conditions, both in BOLD fMRI (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Fox, Spreng, Ellamil, Andrews-Hanna, & Christoff, 2015; Mason et al., 2007) and in M/EEG measures such as amplitude of alpha band activity (Chen, Feng, Zhao, Yin, & Wang, 2008; Compton, Gearinger, & Wild, 2019; Gruberger, Simon, Levkovitz, Zangen, & Hendler, 2011; Marino, Arcara, Porcaro, & Mantini, 2019).

Previous studies have investigated the role of both networks, comparing the neural state between mind wandering and task focused (or other baseline) conditions by analyzing a set epoch before self-reported or probe-caught mind wandering, e.g., (Compton et al., 2019; Jin, Borst, & van Vugt, 2019; Kirschner, Kam, Handy, & Ward, 2012; Mason et al., 2007). We will use a similar technique, to address the dynamic shift between mind wandering and task focused states in function of the underlying neurodynamics of GDN and DMN. Overall, GDN activity anticorrelates with that of DMN (Fox et al., 2005; Fransson, 2005). DMN activity facilitates spontaneous thought generation while suppressing cognitive control whereas, vice versa, GDN activity facilitates cognitive control while suppressing spontaneous thought generation (Luca, Andrew, Alex, & Jason, 2013). When DMN and GDN compete for dominance, episodes of proneness to mind wandering alternate with more focused ones, as one system temporarily gets the upper hand over the other (Figure 1a). The dynamic equilibrium between these neural effects could therefore determine the upcoming mental state.

Simple competition will not suffice, however, to account for the complexity and long-term regularity in the temporal patterns of mind wandering. In Nakatani et al. (2019), experience sampling techniques were used to collect mind wandering reports during the wakeful hours over periods of 14 days. Participants concurrently reported the quantity and experiential content of spontaneous thoughts, and the level of cognitive control was measured by a simple online task. Results were plotted as trajectories in a 2-dimensional space, with the dimensions spontaneous thought and cognitive control (Figure 1b). Regions in this space denote a variety of mental states, for which DMN and GDN activity have been considered as major neural contributors (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016; Voss & Voss, 2014).

As expected, mind wandering occurred in the region where thought production is high and cognitive control is low. However, the trajectory leading to that state was far more complex than what a simple antagonism of DMN and GDN could provide. This would have only allowed movement back and forth between states of low thought production/high control and high thought production/low control, in other words, along the antidiagonal of Figure 1b. Yet, in particular before entering the mind wandering state, the trajectory typically passes through regions in which cognitive control and thought production are both high. Mental states in such regions might, for instance, reflect multi-tasking behavior, as was independently postulated for creative thinking (Fink, Graif, & Neubauer, 2009; Jaarsveld & Van Leeuwen, 2005). The very occurrence of such states flouts a strict antagonism between DMN and GDN effects.

Complex trajectories encompassing such states are still possible when the principle of antagonism applies in a mitigated form. In the DMN, when task focus changes from external stimulation to an internal focus, e.g., autobiographical thoughts, some DMN regions change from task-negative to task-positive (Andrews-Hanna, Saxe, & Yarkoni, 2014; Andrews-Hanna, Smallwood, & Spreng, 2014; Buckner & Dinicola, 2019). Graded shifts are possible as well, as DMN subregions respond differently from one another. This mitigates the antagonism with GDN and allows an enriched repertoire of mental states and dynamic trajectories. In particular, it may help explain the trajectory from a task-focused to a mind wandering state passing through multi-tasking states, as observed in Nakatani et al. (2019). As DMN activity rises, so does spontaneous thought generation. A corresponding downturn in cognitive control activity, however, might be prevented, at least temporarily, by some areas of DMN shifting to task-positive. We will investigate if such a mechanism plays a role in the transition from task-focused to mind wandering states.

Another factor mitigating the antagonism of DMN and GDN may be time lags between neural activity and behavioral effects. Even though delays are known to enable complex dynamic trajectories, most studies have exclusively focused on short-term relations between neural events and behavioral effects at the within-trial level (e.g., using event-related potentials). Neural effects with a latency of multiple trials could have non-trivial effects on transitions between mental states. This mechanism, too, could explain the

transitions from task-focused to mind wandering states in Figure 1. A rise in DMN activity, for instance, could increase thought generation before cognitive control starts weakening, as a result of which the trajectory leading to mind wandering passes through a state in which both thought generation and cognitive control are high.

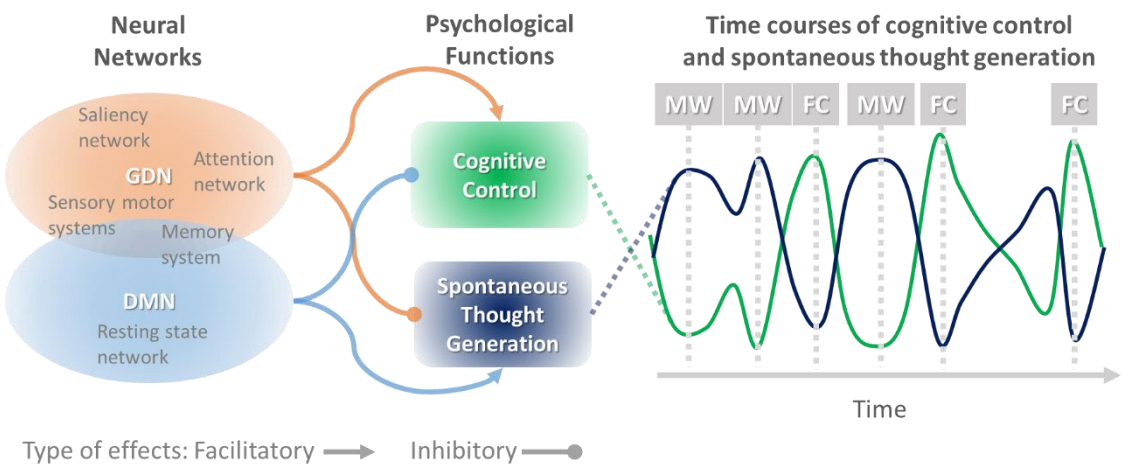
Shifts from task-negative to positive and delayed effects in the relationship between neural activity and psychological functions may be described in terms of a dynamical response function $F_{N,\Psi,t}(\delta t)$, where N is a neural system, {DMN, GDN} and Ψ the psychological response of interest {thought generation, cognitive control}, t is time, and δt is the time lag between neural state and psychological response Ψ (Figure 1c). We will consider measures of $F_{N,\Psi,t}(\delta t)$ as linear predictors of behavior. Predictors at different lags dt will be included, in order to capture delayed effects. The empirical correlates of $F_{N,\Psi,t}(\delta t)$, viz. $\hat{F}_{N,\Psi,S}(\Delta t)$, will be sampled at discrete lags Δt (as the data are sampled over discrete intervals) for N = {DMN-related activity, GDN-related activity} and Ψ = {self-reported thought generation, behavioral measure of cognitive control} in mental states, S = { mind wandering, task focused}. Subscript S is used instead of t to distinguish state-specific effects from general effects of time on task, e.g., fatigue. Instead of a fixed sign of effect (always positive for DMN effects on spontaneous thought generation and GDN effects on cognitive control, and always negative for DMN effects on cognitive control and GDN effects on spontaneous thought generation) the response function can have a positive or negative effect.

In our experiment, participants performed a simple and repetitive, mind-wandering inducing task. Meanwhile, we obtained their S and Ψ from self-report and task performance data. EEG was recorded, from which measures N of DMN and GDN activity were computed. EEG was the chosen method because of its transferability to non-laboratory environments, for which we consider our methods potentially suitable.

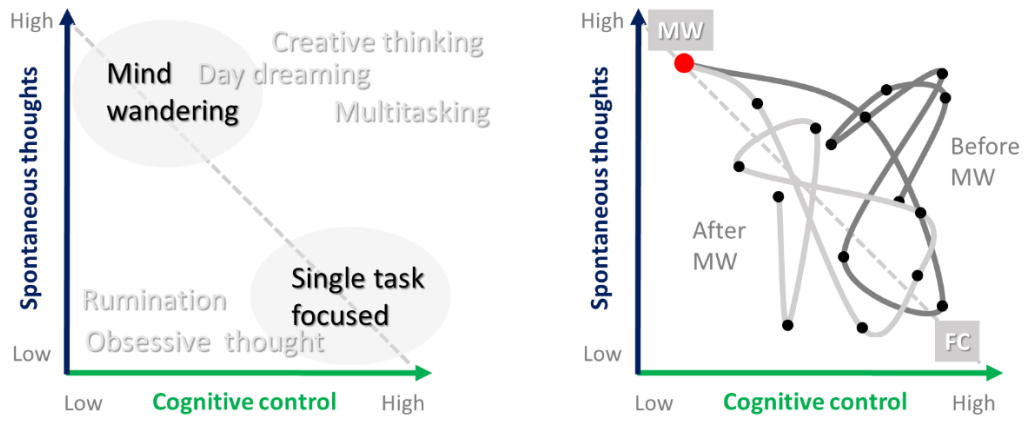
Firstly, we tested whether the measures of DMN and GDN activity could predict future states of the psychological measures. Secondly, we checked the sign of the predictor effect. According to the first mitigation on the antagonism between DMN and GDN, the DMN predictor of the cognitive control measure is to switch in sign from negative to positive as the mental state shifts from task focused to mind wandering. (Andrews-Hanna et al., 2014; Buckner & Dinicola, 2019). Previous studies however did not investigate the

effect of DMN on the spontaneous thought generation measure. Given that only a subset of DMN regions would switch the effect sign, we might expect the DMN effect on spontaneous thought generation to stay positive (i.e., facilitatory for the thought generation). As for the effects of activity, no changes in sign were expected. Thirdly, we consider the role of lags in the switch from focus states to mind-wandering, as the second mitigating factor in the antagonism of DMN and TPN.

a)


Neural Networks
Psychological Functions
Time courses of cognitive control and spontaneous thought generation
Saliency network
GDN
Attention network
Sensory motor systems
Memory system
DMN
Resting state network
Cognitive Control
Spontaneous Thought Generation
MW
MW
FC
MW
FC
FC
Time
Type of effects: Facilitatory
Inhibitory

b)


High
Spontaneous thoughts
Low
Mind wandering
Day dreaming
Creative thinking
Multitasking
Rumination
Obsessive thought
Single task focused
Low
Cognitive control
High
High
MW
Spontaneous thoughts
Low
Before MW
After MW
FC
Low
Cognitive control
High

c)

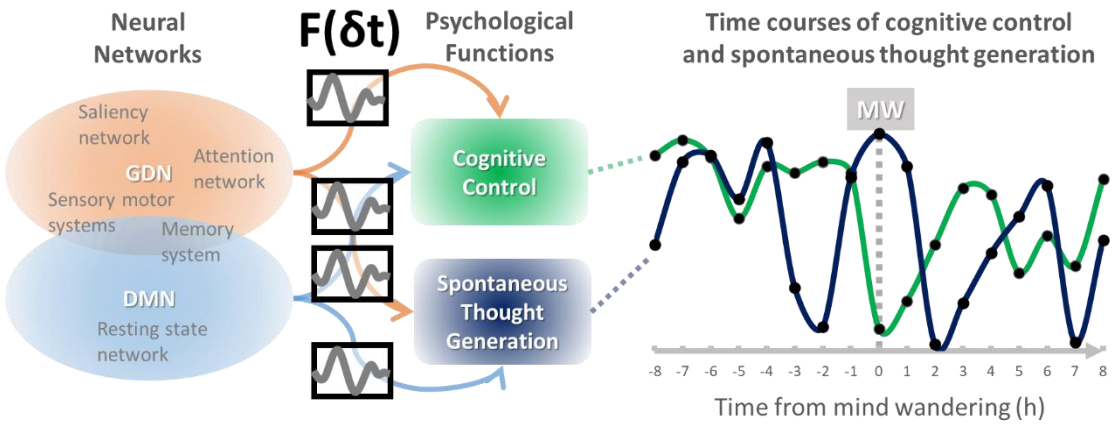

Neural Networks
F(δt)
Psychological Functions
Time courses of cognitive control and spontaneous thought generation
Saliency network
GDN
Attention network
Sensory motor systems
Memory system
DMN
Resting state network
Cognitive Control
Spontaneous Thought Generation
MW
-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8
Time from mind wandering (h)

Figure 1. The classical neurodynamic model, the state space, and the current model for mind wandering and related mental states.
a) In a classical model of goal-directed behavior, goal-directed networks (GDN) collectively facilitate cognitive control, while suppressing spontaneous thought generation. Networks collectively known as the default mode network (DMN) facilitate thought generation, while suppressing cognitive control. The model predicts that around a report of mind wandering (MW) cognitive control (green line) is low and thought generation (dark blue line) is high, while the opposite is the case when a task-focused state (FC) is reported. However, the model fails to take into account possible switches from task-negative to task-positive in the DMN as well as variable time lags between neural activity and behavioral effects.
b) Left panel: Mind wandering and single task focused behavior in a state space defined by the dimensions of spontaneous thought generation and cognitive control. According to the classical model, transitions from the task-focused to the mind wandering state, and vice versa, would gravitate along the antidiagonal (dotted grey line). Right panel: empirical trajectory before (dark gray, 8 hours) and after (light gray, 8 hours) a report of mind wandering (Nakatani et al., 2019). Note the complexity of the trajectory and the divergence from the antidiagonal.
c) DMN and GDN effects on each psychological function as a dynamical response function, $F(\delta t)$ (additional explanation in the text).

Methods

Overview

Participants performed the task of counting tones up to a target number pre-specified at the start of the trial. Each time after reaching the target, they retrospectively rated their mental states during the trial on a scale, ranging from fully task focused to fully mind wandering (%MW), and provided a subjective report of the number of thoughts (NoTs) that had occurred during the trial (Figure 2, top). NoTs were used as measure of spontaneous thought generation, while correct/error of the tone count (correct=1, error=0, %CC stands for the percentage of Correct Count) was used as behavioral measure of cognitive control. EEG was recorded continuously during the task. As the neural measure of GDN activity, amplitude of the auditory evoked potential (AEP) to the tone stimuli was used. AEP amplitude correlates positively with auditory attention, for which the GDN is held responsible (Picton & Hillyard, 1974; Picton, Hillyard, Krausz, & Galambos, 1974). As the neural measure of DMN activity, amplitude of tau rhythm related activity was used. The tau rhythm is an alpha-band (8-12Hz) activity which decreases amplitude from resting to auditory task states (Lehtelä, Salmelin, & Hari, 1997; Tiihonen et al., 1991; Weisz, Hartmann, Müller, Lorenz, & Obleser, 2011). Similar to other resting-state alpha-band activities, such as posterior alpha (visual) and myu rhythm (motor) rhythms, the tau rhythm has been related to DMN activity. Alpha-band activity does not overlap with the main band of the AEP which is situated in the theta band (4-8Hz, Bernhard, 2018).

The 5 single-trial measures (%MW, NoTs, %CC, alpha, and AEP amplitude) were preprocessed as time series per experimental session (Figure 2, middle). The mental state shifts between task-focused and mind wandering states in irregular time intervals during the task. To secure a sufficient length of data for analysis over the various interval of the shifts, we applied event time locked averaging; the time series were segmented from the time points of high %MW reports (≧ 70%) or those of low %MW reports (< 30%) and averaged over the segments per participant.

To the event time locked averages, a multi-variate linear regression model was fit to estimate Alpha and AEP effects to NoTs and %CC. The model was also used to test stochastic causality. Stochastic causality refers to temporal dependence between two time series, which is stochastically reliable enough to

attribute a causal relationship from one to the other (Granger, 1969; Pearl, 2000). The causality is judged based on stricter criteria than those for simple correlation. The test for causality also informs us about the timing of neural effects on the psychological functions. Thus, we tested the causality of the Alpha and AEP effects on NoTs and %CC. For stochastically causal effects, we checked whether the sign changed between the high- and low %MW segments.

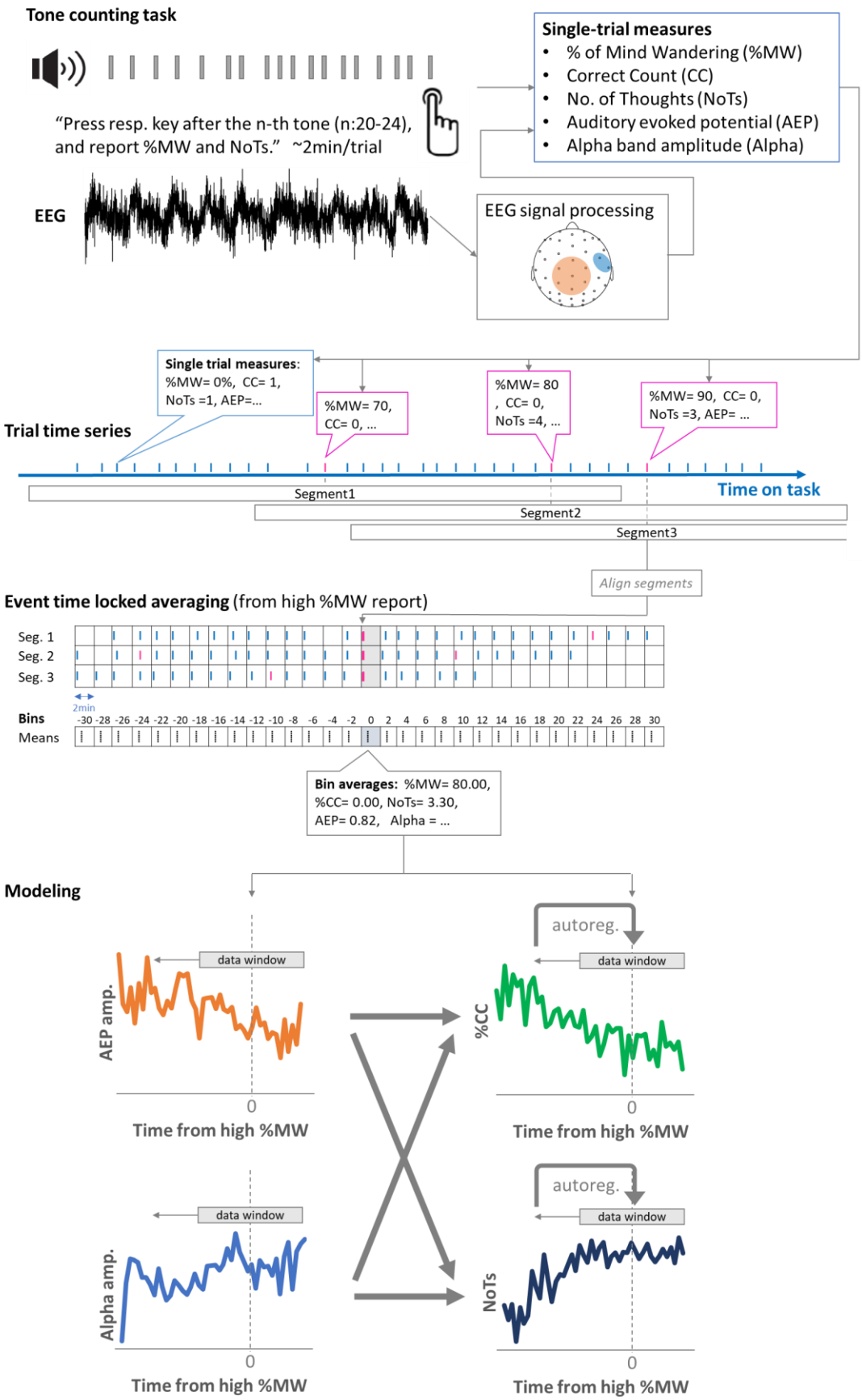

Tone counting task
"Press resp. key after the n-th tone (n:20-24), and report %MW and NoTs." ~2min/trial
EEG
Single-trial measures
% of Mind Wandering (%MW)
Correct Count (CC)
No. of Thoughts (NoTs)
Auditory evoked potential (AEP)
Alpha band amplitude (Alpha)
EEG signal processing
Trial time series
Single trial measures: %MW= 0%, CC= 1, NoTs =1, AEP=...
%MW= 70, CC= 0, ...
%MW= 80 , CC= 0, NoTs =4, ...
%MW= 90, CC= 0, NoTs =3, AEP= ...
Time on task
Segment1
Segment2
Segment3
Align segments
Event time locked averaging (from high %MW report)
Seg. 1
Seg. 2
Seg. 3
2min
Bins
Means
-30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Bin averages: %MW= 80.00, %CC= 0.00, NoTs= 3.30, AEP= 0.82, Alpha = ...
Modeling
data window
AEP amp.
Time from high %MW
autoreg.
%CC
Alpha amp.
NoTs
0

Figure 2. Illustration of the methods
Top, Tone counting task: Participants counted n (20-24) tones, and pressed a key when the count was reached. After the key press, they reported the percentage of time spent for mind wandering (%MW) and the number of thoughts (NoTs) during the trial. EEG was recorded during the task. Amplitude of alpha activity and auditory evoked response (AEP) were computed from the EEG. Therefore, each trial yields 5 measures, Correctness of the count (CC), NoTs, %MW, Alpha, and AEP amplitude.
Middle, Event time locked averaging: From the moment of high % mind wandering report (%MW ≧ 70%) data segment was defined. The segments were then aligned and averaged. (The same procedure was also applied to the low %MW report.)
Bottom, Illustration of main data analysis: The model was applied to estimate stochastic causality, direction and timing of the neural effects.

Ethical approval and Participants

The study was approved by The KU Leuven Medical Ethics committee (protocol number: S56128). Twenty healthy volunteers (average 23.00 years, 6 male) participated in exchange for monetary compensation (10€/hour). They were recruited via the online participant recruitment system at KU Leuven. Prior to the participation, they were given information about the study and gave oral and written informed consent. All data were collected prior to the onset of the 2020 COVID19 pandemic.

Tasks

The *tone counting* task (Figure 2, top) was used to elicit a sizeable incidence rate of mind wandering during an approx. 2-hour period. The task consisted of 37 trials lasting approx. 90s each, in which a sequence of brief tones, 100ms in 440Hz frequency, was presented at 70dB via the speakers located at right and left sides of a monitor in front of them. The tones were presented with a stimulus onset asynchrony (SOA) between 2000 and 2500ms (pseudo-randomized with 100ms steps). At the beginning of each trial, a number was presented in the center of the display. The number indicated how many tones had to be counted before a response. The number of tones to be counted ranged between 20 and 24 (pseudo-randomized over trials). Participants put the right index finger on Key "5" of the ten-key pad and closed their eyes before the series of tones started. When the target number was reached, participants responded by pressing the key. Then they opened their eyes and reported the percentage of time spent on task irrelevant thoughts in 0 to 100% in 10% steps. An eleven-point scale (0% --10%--....-90%--100%) was shown on the display and participants reported the percentage (%MW) using the ten-key pad. Lastly, participants reported on the keypad the number of thoughts (NoTs) that had occurred in their minds during the trial on a ten-point scale. The scale ranged from 1: only task-related thought to 10: task-related and 9 or more task-irrelevant thoughts. Participants were instructed to evaluate the number of thoughts in a consistent manner over trials. For instance, they were instructed to report "1" when they had only task-related thoughts; when they had one task-irrelevant thought, for instance about having lunch after the experiment, they reported "2". They were told that items included in the thoughts (e.g., a sandwich, an apple, and a bottle of soda) were not counted separately. If two different thoughts occurred at different times during the trial (e.g., thoughts on lunch,

then later during the trial thinking about doing laundry), they reported "3." And so on. The instruction was the same as that in Nakatani et al., (2019). Participants had the opportunity to take a break at marked sections of the task. EEG was recorded during the task.

An *Auditory Oddball Task* was applied prior to the tone-counting task to specify the relevant parameters of the auditory evoked potentials (AEP), namely, scalp distribution, latency, and main frequency band, for subsequent use in the tone counting task. Participants placed their right index finger on the "5" key of the ten-key pad and closed their eyes. Then, a series of 110 tones was presented at 70 dB via the speakers, with an SOA pseudo-randomized between 2000 and 2500ms in 100ms steps. Of these tones, 101 were standard (440Hz, 100ms duration), physically identical to the ones used in the tone counting task, and 5 were odd-ball tones (510Hz, 100ms duration), which were randomly interspersed. Participants were asked to respond to the oddballs by pressing the key as fast as possible.

Both tasks were implemented using PsychoPy Version 1.82.01 (Peirce, 2009) on Windows XP.

EEG recording

EEG was recorded using the Patientguard EEG recording system (ANT Neuro, Enschede, The Netherlands). Of 64 channels in total, 59 were EEG channels and the remaining ones were used for recording the electrooculogram (EOG, vertical and horizontal) and recording reference (right mastoid, was later re-referenced to average reference). The placement of the EEG electrodes was according to the manufacturer-defined whole head system rather than the 10-10 system. The signal was sampled at 2048Hz. Electrode impedance was 20 kOhm or less (for the majority, the impedance was 5 kOhm or less).

Procedure

Experiments were conducted with participants individually in an EEG-laboratory of the University Hospital of Leuven. After electrodes were attached, the participant was seated in a dimly lit room in front of speakers and a monitor. Baseline EEG was recorded (eyes open/relax 1 min, and eyes close/relax, 1 min). This was followed by an auditory oddball task for estimation of evoked response parameters, and next by the tone counting task. EEG was recorded during both tasks. Afterwards, baseline EEG measurement was performed

again. Then the electrodes were removed, and the participant washed off the gel before leaving. The experimental session took about 3 hours in total.

Preprocessing for subjective reports and behavioral data

Event time locked averaging was applied to time series of trial data (Figure 2, middle). Trials with %MW≧70 and those with %MW<30 were labeled as mind wandering (MW) and task-focused (FC) trials, respectively. The thresholds, 70% and 30%, were decided based on histogram of %MW (Appendix Fig. A1). The moment of the high (or low) %MW report was marked in time-on-task as Time0 for segmentation. From Time0, time series were segmented between -50min and +32min. The segment was extended to the past to estimate autoregressive effects (See Models for more details). On average 9.11 and 6.53 segments/participant were obtained for the high- and low %MW reports, respectively. In both types of report, segments were aligned from Time0 for averaging. However, trials do not necessarily align in time (except for those including Time0 of a given high- and low %MW reports), since the inter-trial interval varied according to the number of tones per trial and to individual differences in report timing. Thus, we set bins of 2-min intervals; the data points between Time0 and 1min 59.99s were averaged and considered as the data for Bin0. Similarly, the next interval was between 2min and 3min 59.99s, and the averages were data for Bin2. Note that the bin numbers correspond to the starting time of the bin, e.g., Bins2, 4, and 6 correspond to bins which start 2, 4, and 6 min after Time0, and Bins-2, -4, and -6 start 2, 4, and 6 min *before* Time 0, respectively. After the averaging the time interval of the data was uniform 2min. For each participant, the interval series was computed for NoTs and %CC. and used for the following analysis.

Preprocessing for EEG data

EEG data were pre-processed using Brain Vision Analyzer (version 2.1.2, Brain Products GmbH, Germany) and in-house Python scripts. Data were down-sampled from 2048 to 1024Hz, re-referenced to the average of all EEG channels, and filtered between 0.32 and 45.00Hz using a 2nd order zero phase shift Butterworth filter. Independent component analysis (ICA, Infomax algorithm) was applied to the filtered data for artifact removal. EOG channels were included in the ICA analysis. Ocular, cardiac, and muscular artifact components were visually identified and excluded from the inverse ICA. After the cleaning, two

channels (one for 2 participants) still showed poor signal quality, thus, they were interpolated using adjacent channels.

Computation of AEP amplitude

Scalp distribution, peak latency, and peak frequency band of AEP were identified using the EEG data of the standard tones in the odd-ball task. The data were segmented from the onsets of the standard tones, between -500 and +1500ms. The complex-Morlet wavelet transform was applied to the segment. Amplitude was computed from the complex-valued analytic signal. The amplitude in the time-frequency domain was normalized and averaged for each participant. The t-statistic and accompanying probability were computed for each time-frequency point. Type-I error was corrected using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995). FDR was applied to the p-values of all data points over the 59 EEG electrodes. Amplitude in the frequency band between 4 and 8Hz, in time between 130 and 430ms from the stimulus onset, and in 7 central electrodes (3Z, 4Z, 5Z, 4L, 5L, 4R, 5R in the WaveGuard Duke64 system, cf. Appendix Fig. A2) showed p(FDR-corrected) < 0.05. The frequency, latency, and scalp distribution matched those in previous AEP studies (Picton & Hillyard, 1974; Picton et al., 1974).

AEP amplitude in the tone counting task was computed based on the parameters specified above; EEG data of the tone counting task were band-passed between 4 and 8Hz using an in-house Python code for band-pass filtering which combines a Butterworth filter (scipy.signal.butter, 12dB) and zero-phase shift filter (scipy.signal.filtfilt). Instantaneous amplitude was computed using the complex-valued analytic signal from the Hilbert transform of the band-passed signal. The amplitude signal was segmented between -500 and 1000ms from a tone onset. Each segment was normalized using the pre-stimulus mean and standard deviation. Then the segments were averaged per trial (over 20-24 segments). Using the average waveforms, amplitude values were averaged between 130 and 430ms, and over the 7 central electrodes. This average was used as the AEP amplitude of the trial.

Computation of tau-related alpha-band amplitude

Previous studies have reported resting-state alpha band activity, tau rhythm, for auditory task paradigms (Lehtelä et al., 1997; Tiihonen et al., 1991; Weisz et al., 2011). Scalp distribution of the activity was

reported as bilateral temporal. The peak frequency slightly differed among the previous studies. Considering the difference and the frequency band used for AEP, EEG was band passed between 8-12Hz (central frequency of 10Hz) using the above mentioned in-house band-pass filter. To the band-passed signal Hilbert transform was applied to compute instantaneous amplitude. The amplitude was averaged for the pre-task eyes-closed resting period using approx. 1 min of noise-free segments. The average amplitude was compared between the rest and the first task trial in all electrodes. The trial had 20-24 tones. From the onset of each stimulus, a period between 0 and 2000ms was averaged and further averaged over the stimuli. The alpha blocking was, however, not statistically reliable in the temporal electrodes. The results may be in part due to the presence of posterior alpha-band activity. The posterior alpha was present during the tone counting task because the participants closed eyes during the task. This visual resting-state alpha activity spreads over large posterior regions, thus could be mixed with tau rhythm in temporal electrodes. Identification of tau rhythm became difficult as a result.

Moreover, the posterior alpha seemed to cause another problem. As we will explain in Models, the main analysis consisted of a time series analysis. As preprocessing, event time locked averaging was applied to the alpha in the same way as for NoTs and %CC. After the averaging, time series of the alpha amplitude became ‘flat’ in many electrodes. Such signals without temporal change cannot be used for the main analysis. The low variance might be due to the prominent posterior alpha. In other words, the small number of unsaturated signals could reflect other alpha band activity more, including tau rhythm. Thus, we checked which channel was not saturated in terms of the alpha amplitude after event time locked averaging. For the checking, data segments different from those for the main analysis were used to avoid circular analysis; data segments from intermediate-level %MW report, $40 \leqq \%MW < 60$, between -30 and +32min were used for event time locked averaging. The segments were aligned form the time marker of the %MW report and averaged, on average over 10.58 segments/participant. To the averaged time series, a mixed effect model, Alpha ~ Time (fixed effect factor) + participants (random effect factor) was fit using the R package, lme4 (Bates, Mächler, Bolker, & Walker, 2014). The effect of Time was estimated using chi-squared (χ^2) statistic of ANOVA between the model and its null model: Alpha ~ participants. The test was applied to all EEG

electrodes. Type-I error over the univariate tests was corrected over the electrodes using the FDR method. Result showed sizable effect of Time in 2 adjacent right temporal electrodes (3RB and 3RC; $\chi^2(1) = 11.23$, p(FDR-corrected) = 0.03, $\eta^2 = 0.43$, power = 0.38 and $\chi^2(1) = 12.33$, p(FDR-corrected) = 0.02, $\eta^2 = 0.44$, power = 0.34, respectively). The signals from the electrodes could be considered as representing tau, since the scalp distribution agrees to that in previous studies. At the same time, the signals could also contain heterogenous alpha subtypes, e.g., posterior/visual resting state activity which is also DMN-related. Thus, we considered the signals as DMN-related, and refer to it as Alpha from hereon. Similar to AEP, amplitude of Alpha was averaged over the 2 channels to make a single measure of DMN-related activity.

Event time locked averaging of AEP and Alpha

Finally, event time locked averaging was applied to the time series of AEP and Alpha (Figure2, middle). In each measure, data segments between -50 and +32min from Time0 was used for the averaging. The procedure of the averaging was the same as that for NoTs and %CC. The averaging generated interval series between Bins-25 to +15 from Bin0, for the high- and low %MW report separately.

Models and main data analysis

Models were built to estimate stochastic causality, timing, and sign of the neural effects. Stochastic causality refers to temporal dependence between two time series, X and Y, which is statistically reliable enough to consider causal relationship from X to Y (Granger, 1969). The current state of Y could be predicted by its own past states. Besides the autoregressive states, if past states of X could improve the prediction of Y's current state, X could be considered as stochastically causal to Y. We build the model step-by-step, starting from autoregression model. For NoTs, the autoregression model can be written as

(Eq.1)

$$NoTs_t = \sum_{i=1}^{N} a_i * NoTs_{t-i} + \varepsilon_t$$

NoTs at a given time point *t*, $NoTs_t$, is predicted by the weighted sum of own past states $NoTs_{t-i}$, where *i* is the index for time regression and a_i is the weight. The ε_t is the error term. The past states were summed over *N*. To simplify the model, we temporarily set *N=1*. The simplified model is

(Eq.2)

$$NoTs_t \approx a_1 * NoTs_{t-1} + \varepsilon_t$$

To the autoregression model, 2 neural predictors, $Alpha_{t-1}$ and AEP_{t-1}, were added as follows.

(Eq.3)

$$NoTs_t \approx a_1 * NoTs_{t-1} + b_1 * Alpha_{t-1} + c_1 * AEP_{t-1} + \varepsilon_t$$

The neural and autoregression predictors were used to test stochastic causality of the neural effects. To the 3-predictor model, two more neural predictors, $Alpha_t$ and AEP_t, were added. The predictors estimate neural effects without time difference from the response variable.

(Eq.4)

$$NoTs_t \approx a_1 * NoTs_{t-1} + (b_1 * Alpha_{t-1} + c_1 * AEP_{t-1})$$
$$+ (b_0 * Alpha_t + c_0 * AEP_t) + d_0 * Participants + \varepsilon_t$$

Parentheses are added for readability. Also, a categorical variable $Participants$ was added to take into account for individual differences. The same formulation was applied to predict %CC at time *t*, $\%CC_t$.

(Eq.5)

$$\%CC_t \approx a_1 * \%CC_{t-1} + (b_1 * Alpha_{t-1} + c_1 * AEP_{t-1})$$

$$+ (b_0 * Alpha_t + c_0 * AEP_t) + d_0 * Participants + \varepsilon_t$$

NoTs and %CC were two parallel data series in the tone counting task which could covary. To account for the covariance, we chose multivariate method to estimate contribution of each predictor. The model for the estimation may be written as pseudo R code.

(Eq.6)

$$[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-1}) + (Alpha_{t-1} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$$

The multivariate linear regression method estimates the contribution of the autoregression, past-neural, present-neural predictors and the individual difference term to the prediction of the present state of NoTs and %CC, taking into account possible correlation among the variables. Reduced models can be made from this model (Eq.6) to evaluate stochastic causality of Alpha and AEP effects; a model without $Alpha_{t-1}$ and another without AEP_{t-1}. If the reduced models predict the response variables worse than the full model did, the contribution of the omitted neural predictor is not negligible. I.e., the comparison gives the basis for stochastic causality for the effect.

The above model set the time regression as *t-1*, which means the most recent past state represents the effect of all past states. This would be appropriate if we could assume that the most recent past state has always the strongest effect. However, as discussed in Introduction, we assumed that the effect timing could change between task-focused and mind wandering states. Thus, the order of regression was parametrized using indices *i, j, k,* and *l*.

(Eq.7)

$$[NoTs_t, \%CC_t] \sim \left(NoTs_{t-i} + \%CC_{t-j}\right) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$$

In the current study, the parameters vary systematically between 1 and 10 (based on the data bin size 2min, this corresponds to time regression between 2 and 20min. The range was decided by the length of the data). The full combination of the 4 parameters gives 10^4 models to fit. The number of models were reduced in 2 steps. In the first step, *i* and *j* was searched between 1 and 10 for the highest autoregressive effects, while *k* and *l* were fixed to *k=1* and *l=1*. The larger autoregression terms were desirable since they make the criterion for stochastic causality stricter. The resulting 10 by 10 = 100 models were fit to the high- and low %MW segments, separately. Upon fitting, a data window was applied; the window for response variables, $NoTs_t$ and $\%CC_t$, was ±15 bins from Bin0 (which starts from *t*, the moment of the %MW report) based on the results of subjective report and task measure in our previous study (Nakatani et al., 2019). The same data window was applied to the present neural predictors. For the autoregression and past neural state predictors, the size of the data window was the same, however, the window was shifted bin-by-bin, up to 10bins (20min) to the past, as indicated by *i* and *j*. Of the 100 combinations of *i* and *j*, those with the highest autoregressive effects were selected using adjusted-R^2. In case of tie, the smallest *i* and *j* were chosen e.g., if (*i=1, j*=2) and (*i=2, j*=2) gave the highest adjusted-R^2, (*i=1, j*=2).

In the second step, *k* and *l* were systematically varied between 1 and 10, while the autoregression terms were fixed with *i=i** and *j=j**, where *i** and *j** were those selected in the first step. In this manner, the number of full models were reduced from 10^4 to 10^2. The full model with *i** and *j** in pseudo R code is

(Eq.8)

$$[NoTs_t, \%CC_t] \sim \left(NoTs_{t-i*} + \%CC_{t-j*}\right) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$$

The models were fit to the high- and low %MW segments, separately. To test the stochastic causality, 2 reduced models, one without the past Alpha and the other without the past AEP term, were also fit to the data, the reduced models were

(Eq.9)

$$[NoTs_t, \%CC_t] \sim \left(NoTs_{t-i*} + \%CC_{t-j*}\right) + AEP_{t-l} + (Alpha_t + AEP_t) + Participants$$

(Eq.10)

$$[NoTs_t, \%CC_t] \sim \left(NoTs_{t-i*} + \%CC_{t-j*}\right) + Alpha_{t-k} + (Alpha_t + AEP_t) + Participants$$

The Pillai test was used to estimate statistical significance of the difference between the fits of the full and reduced models. The 100 comparisons for 2 reduced models yielded 200 tests. The R implementation of Pillai test provides the estimated F-statistic with corresponding probability. We corrected for Type-I errors over the 200 tests by using the FDR method. An arbitrary threshold ($p<0.01$) was applied to the corrected probability, to attribute stochastic causality to the neural effect. The indices k and l reveal the timing of the effect, e.g., a stochastically causal effect of $Alpha_{t-k}$ at $k=4$ to $NoTs_t$ suggests that (given the 2min interval of data bin) the neural activity between 6-8min *before* the report of the high (or low) %MW had stochastically causal consequence to the number of spontaneous thoughts.

Results

Prior to the model fit, multi-variate outliers were removed using Mahalanobis distance among NoTs, %CC, Alpha and AEP data; datapoints with Mahalanobis distance > 20 were excluded from the model fit. Also, the linear trend was removed. The model was fit to the average high- and low %MW data segments, separately.

Results for High %MW segments

The first step of model fitting suggested *i = 2* and *j = 5* for the highest autoregression effect (See *Notes on autoregressive effects* for more details). For the second step, *i** and *j** were set as 2 and 5, respectively, and 100 models were generated over the combination of *k* and *l*. The full- and corresponding reduced models were fit to the high %MW segments data. The difference between the full- and reduced models was tested using the Pillai test. Type-I error was corrected over the 200 tests using FDA method. Table1 shows the corrected probability of the estimated-F statistic of the Pillai score. With an arbitrary threshold of p<0.01, stochastic causality was suggested for past neural predictor $Alpha_{t-k}$ at k = 4 and 5 (Table 1a), while no causality was suggested for AEP_{t-l} (Table 1b). Timing of the effect at k = 4 and 5 correspond to 6-8 and 8-10min before the high %MW report, respectively.

Tables 2 list the estimated coefficients for the full model with *k=4 and 5*. (For both cases, the index for AEP_{t-l} was *l=1*. With the choice, no crucial information was missed since the effect of AEP_{t-l} changed little over *l*.) The coefficient for the past neural effect $Alpha_{t-4}$ which was positive as a predictor for $\%CC_t$; coefficient = 8.16, the 95% Confidence Interval [2.5%, 97.5%] = [4.22, 12.11], t-value = 4.07. The effect was sizable compared with that of the autoregression term, $\%CC_{t-5}$; coefficient = 0.21, 95%CI = [0.12, 0.31], t-value = 4.33. By contrast, the effects of the present neural predictors $Alpha_t$ and AEP_t to $\%CC_t$ were negligible. For the prediction of $NoTs_t$, no neural effect was noteworthy, while the autoregression effect was considerable; coefficient of $NoTs_{t-2}$ = 0.24, 95%CI = [0.15, 0.33], t = 5.29. Results of *k=5* (Table 2b) showed the same pattern as *k=4*; The effect of $Alpha_{t-5}$ to $\%CC_t$ was positive, and also the auto regression effects were positive.

The above results show that Alpha activity 6-10min before the high %MW report had a stochastically causal and task-*positive* effect to the %CC. The rest of neural effects did not show notable contribution to predict the %CC and NoTs.

Table 1. Effect of past neural states in High %MW condition.
a) Model comparison tests the stochastic causality of past neural Alpha states, b) tests the stochastic causality of past neural AEP states.
Pillai test was applied to test the difference between full and reduced models. Probability of estimated F-statistic, DF = (1, 462) of Pillai score was shown after FDR correction.

a) $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$
vs. $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + AEP_{t-l} + (Alpha_t + AEP_t) + Participants$

		AEP $_l$									
		l=1	l=2	l=3	l=4	l=5	l=6	l=7	l=8	l=9	l=10
Alpha $_k$	k=1	0.0649	0.0649	0.0649	0.0649	0.0649	0.0649	0.0649	0.0649	0.0651	0.0649
	k=2	0.0335	0.0339	0.0335	0.0337	0.0335	0.0326	0.0337	0.0326	0.0335	0.0335
	k=3	0.2282	0.2282	0.2282	0.2282	0.2300	0.2282	0.2282	0.2282	0.2301	0.2282
	k=4	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0009	0.0005	0.0005	0.0005
	k=5	0.0025	0.0025	0.0025	0.0025	0.0025	0.0025	0.0031	0.0027	0.0025	0.0025
	k=6	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2277	0.2282
	k=7	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282	0.2282
	k=8	0.2318	0.2676	0.2282	0.2301	0.2301	0.2402	0.2402	0.2634	0.2301	0.2282
	k=9	0.5072	0.5071	0.5102	0.5071	0.5105	0.5071	0.5197	0.5071	0.5071	0.5071
	k=10	0.2301	0.2301	0.2301	0.2301	0.2372	0.2301	0.2634	0.2301	0.2402	0.2301

p<0.01 p<0.001

b) $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$
vs. $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + Alpha_{t-k} + (Alpha_t + AEP_t) + Participants$

		AEP $_l$									
		l=1	l=2	l=3	l=4	l=5	l=6	l=7	l=8	l=9	l=10
Alpha $_k$	k=1	0.5298	0.0649	0.5071	0.3540	0.9296	0.0649	0.3275	0.5071	0.5735	0.5071
	k=2	0.5551	0.0649	0.5072	0.4219	0.9709	0.0649	0.3479	0.4721	0.5210	0.5071
	k=3	0.5626	0.0649	0.5071	0.3741	0.9709	0.0649	0.2634	0.5071	0.5565	0.5071
	k=4	0.5564	0.0649	0.5071	0.3640	0.9935	0.0649	0.5071	0.4869	0.5071	0.5071
	k=5	0.5626	0.0649	0.5071	0.3741	0.9945	0.0649	0.4219	0.5197	0.5565	0.5071
	k=6	0.5626	0.0649	0.5071	0.3741	0.9629	0.0651	0.3453	0.5071	0.5071	0.5071
	k=7	0.5298	0.0649	0.5071	0.3595	0.9629	0.0649	0.3479	0.5071	0.5071	0.5071
	k=8	0.6141	0.0655	0.5071	0.3759	0.9629	0.0764	0.3595	0.5490	0.5298	0.5071
	k=9	0.5629	0.0649	0.5072	0.3670	0.9629	0.0649	0.3312	0.5071	0.5071	0.5071
	k=10	0.5626	0.0649	0.5071	0.3741	0.9709	0.0651	0.3622	0.5071	0.5551	0.5071

Table 2 Estimated coefficients of the full model for High %MW segments, separately for $Alpha_{t-4}$ (a) and $Alpha_{t-5}$ (b).

a) $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + (Alpha_{t-4} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC** $_t$	Intercept	-4.39	-7.06	-1.72	-3.23	0.001
	NoTs $_{t-2}$	-0.37	-3.81	3.07	-0.21	0.834
	%CC $_{t-5}$	0.21	0.12	0.31	4.33	0.000
	Alpha $_{t-4}$	8.17	4.22	12.11	4.07	0.000
	AEP $_{t-1}$	0.93	-5.01	6.87	0.31	0.760
	Alpha $_t$	-0.44	-4.16	3.28	-0.23	0.815
	AEP $_t$	4.78	-1.07	10.64	1.61	0.109
	Participants	0.39	0.17	0.61	3.46	0.001

Response var. = **NoTs** $_t$	Intercept	-0.02	-0.09	0.05	-0.48	0.631
	NoTs $_{t-2}$	0.24	0.15	0.33	5.29	0.000
	%CC $_{t-5}$	0.00	0.00	0.00	0.54	0.588
	Alpha $_{t-4}$	0.09	-0.01	0.19	1.71	0.088
	AEP $_{t-1}$	-0.09	-0.24	0.06	-1.17	0.244
	Alpha $_t$	0.07	-0.03	0.16	1.39	0.164
	AEP $_t$	0.11	-0.04	0.27	1.49	0.137
	Participants	0.00	0.00	0.01	1.38	0.169

b) $[NoTs_t, \%CC_t] \sim (NoTs_{t-2} + \%CC_{t-5}) + (Alpha_{t-5} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC $_t$**	Intercept	-4.48	-7.16	-1.80	-3.28	0.001
	NoTs $_{t-2}$	-0.60	-4.06	2.86	-0.34	0.732
	%CC $_{t-5}$	0.20	0.11	0.30	4.06	0.000
	Alpha $_{t-5}$	6.92	3.00	10.85	3.46	0.001
	AEP $_{t-1}$	2.35	-3.63	8.32	0.77	0.440
	Alpha $_t$	0.46	-3.22	4.14	0.25	0.807
	AEP $_t$	3.48	-2.40	9.37	1.16	0.246
	Participants	0.40	0.18	0.62	3.55	0.000

Response var. = **NoTs $_t$**	Intercept	-0.02	-0.09	0.05	-0.50	0.619
	NoTs $_{t-2}$	0.24	0.15	0.33	5.23	0.000
	%CC $_{t-5}$	0.00	0.00	0.00	0.40	0.688
	Alpha $_{t-5}$	0.09	-0.01	0.19	1.76	0.079
	AEP $_{t-1}$	-0.07	-0.23	0.08	-0.95	0.343
	Alpha $_t$	0.08	-0.02	0.17	1.59	0.113
	AEP $_t$	0.10	-0.05	0.25	1.29	0.198
	Participants	0.00	0.00	0.01	1.41	0.159

Results for Low %MW segments

Results of the first-stage model fit suggested *i= 1* and *j= 9* for the highest autoregression effect. Thus, *i** and *j** of the model in the second stage was set as 1 and 9, respectively. With the selected autoregression terms, 100 models were generated over the combination of *k* and *l* between 1 and 10. The full- and reduced models were fit to the data in the low %MW segment. The Pillai test was applied in the same way for the analysis of high %MW segment data, to assess the stochastic causality of the neural effects. As shown in Table 3, the test results suggested the causality for $Alpha_{t-k}$ at k = 3, while no causality was suggested for AEP_{t-l} . The corresponding timing for $Alpha_{t-3}$ was 4-6min before the low %MW report.

Table 4 lists the estimated coefficients for the predictors at k = 3. Similar to Table 2, index *l* for AEP_{t-l} was set to *1* to simplify the presentation of the results. The coefficient for the past neural effect term $Alpha_{t-3}$ for $\%CC_t$ was large; coefficient = -10.04, 95% CI = [-14.6, -5.43], t-value = -4.29, relative to that of autoregression term, $\%CC_{-9}$, coefficient = 0.15, 95%CI = [0.07, 0.24], t-value = 3.51. Thus, the past neural effect could be qualified for stochastic causality similar to that in the high %MW segment, although the sign was *negative*. Moreover, in the low %MW segments, the effect of present Alpha was noticeable; coefficient of $Alpha_t$ = 7.38, 95% CI = [2.77, 12.00], t-value = 3.15.

In contrast, $Alpha_{t-3}$ predicted little of $NoTs_t$; coefficient = 0.10, 95%CI = [0.00, 0.21], t-value = 1.89, relative to that of the autoregressive term $NoTs_{-1}$ coefficient = 0.12, 95%CI = [0.03, 0.21], t-value = 2.57. Besides, the effect of present Alpha was noticeable; coefficient of $Alpha_t$ = -0.17, 95% CI = [-0.28, -0.07], t-value = -3.22.

The above results show that Alpha activity 4-6min before the low %MW report had a stochastically causal and task-*negative* effect on %CC. Moreover, the results show that the Alpha activity correlated without delay (in the current time scale), positively with %CC, while negatively with NoTs in the low %MW segment.

Table 3. Effect of past neural states in Low %MW condition.
a) Model comparison tests the stochastic causality of past neural Alpha states, b) tests the stochastic causality of past neural AEP states.
Pillai test was applied to test the difference between full and reduced models. Probability of estimated F-statistic, DF = (1, 395) of Pillai score was shown after FDR correction.

a) $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-9}) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$
vs. $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-9}) + AEP_{t-l} + (Alpha_t + AEP_t) + Participants$

		AEP $_l$									
		l=1	l=2	l=3	l=4	l=5	l=6	l=7	l=8	l=9	l=10
Alpha $_k$	k=1	0.9552	0.9552	0.9552	0.9629	0.9552	0.9552	0.9552	0.9552	0.9935	0.9552
	k=2	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935
	k=3	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014
	k=4	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935
	k=5	0.0952	0.0952	0.0988	0.0952	0.0952	0.0952	0.0952	0.0952	0.0952	0.0952
	k=6	0.3069	0.3222	0.3222	0.2485	0.3038	0.3069	0.3069	0.3222	0.2898	0.3069
	k=7	0.5537	0.5579	0.5719	0.6140	0.5683	0.5579	0.5579	0.5286	0.5579	0.5579
	k=8	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935	0.9935
	k=9	0.0636	0.0453	0.0636	0.0636	0.0636	0.0636	0.0636	0.0928	0.0636	0.0636
	k=10	0.7600	0.7526	0.8904	0.7585	0.7585	0.7492	0.7585	0.7600	0.7492	0.7585

p<0.01

b) $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-9}) + (Alpha_{t-k} + AEP_{t-l}) + (Alpha_t + AEP_t) + Participants$
vs. $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-9}) + Alpha_{t-k} + (Alpha_t + AEP_t) + Participants$

		AEP $_l$									
		l=1	l=2	l=3	l=4	l=5	l=6	l=7	l=8	l=9	l=10
Alpha$_p$	k=1	0.9935	0.7492	0.6698	0.7492	0.9935	0.7585	0.9935	0.4371	0.1102	0.9935
	k=2	0.9935	0.7492	0.6698	0.7492	0.9935	0.7585	0.9935	0.4371	0.0952	0.9935
	k=3	0.9935	0.7492	0.6698	0.7492	0.9935	0.7600	0.9935	0.4435	0.1077	0.9935
	k=4	0.9935	0.7492	0.6698	0.7492	0.9935	0.7585	0.9935	0.4371	0.0952	0.9935
	k=5	0.9935	0.7492	0.7492	0.7492	0.9935	0.7803	0.9935	0.4535	0.0952	0.9935
	k=6	0.9935	0.7585	0.7296	0.6835	0.9935	0.7585	0.9935	0.4562	0.0952	0.9935
	k=7	0.9935	0.7492	0.6914	0.7585	0.9935	0.7585	0.9935	0.4099	0.0952	0.9935
	k=8	0.9935	0.7492	0.6698	0.7492	0.9935	0.7585	0.9935	0.4371	0.0952	0.9935
	k=9	0.9935	0.5083	0.6698	0.7526	0.9552	0.7492	0.9935	0.5683	0.0952	0.9935
	k=10	0.9935	0.7492	0.7492	0.7492	0.9935	0.7492	0.9935	0.4392	0.0952	0.9935

Table 4. Estimated coefficients of the full model for Low %MW condition for $Alpha_{t-3}$.
$[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-9}) + (Alpha_{t-3} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC** $_t$	Intercept	-6.54	-10.27	-2.82	-3.45	0.0006
	NoTs $_{t-1}$	4.83	0.81	8.86	2.36	0.0188
	%CC $_{t-9}$	0.15	0.07	0.24	3.51	0.0005
	Alpha $_{t-3}$	-10.04	-14.64	-5.43	-4.29	0.0000
	AEP $_{t-1}$	1.85	-4.59	8.30	0.57	0.5721
	Alpha $_t$	7.38	2.77	12.00	3.15	0.0018
	AEP $_t$	-1.01	-7.61	5.58	-0.30	0.7628
	Participants	0.57	0.26	0.88	3.65	0.0003

Response var. = **NoTs** $_t$	(Intercept)	-0.07	-0.15	0.02	-1.50	0.1343
	NoTs $_{t-1}$	0.12	0.03	0.21	2.57	0.0107
	%CC $_{t-9}$	0.00	0.00	0.00	-0.27	0.7887
	Alpha $_{t-3}$	0.10	0.00	0.21	1.89	0.0592
	AEP $_{t-1}$	0.03	-0.12	0.18	0.37	0.7105
	Alpha $_t$	-0.17	-0.28	-0.07	-3.22	0.0014
	AEP $_t$	0.04	-0.11	0.19	0.51	0.6115
	Participants	0.01	0.00	0.01	2.06	0.0401

Notes on autoregressive effects

The indices for autoregression terms, $i*$ and $j*$, were selected to make the criterion for stochastic causality as strict as possible, and to reduce the number of full models. For this purpose, we chose *i* with the highest adjusted-R^2 for NoTs, and *j* with the highest adjusted-R^2 for %CC in the first step. The selected $i*$ and $j*$ were 2 and 5 for the high-, and 1 and 9 for the low %MW segments. The timing of the effects, in particular those for the autoregression effect of %CC, were further in the past than might have been expected. As a reality check, we also fit full models with autoregression terms with *i*= 1 and *j*=1 to the data. The results showed that the contribution of the terms were slightly less than those with $i*$ and $j*$; e.g., $\%CC_{t-1}$ had coefficient= 0.11, 95%CI=[0.01, 0.21], t=2.17 (prob.=0.031), in comparison with $\%CC_{t-9}$ coefficient = 0.15, 95%CI=[0.07, 0.24], t=3.51 (prob.=0.0005) for the low %MW segment (See Appendix 3 for more results). The observations suggest that the $i*$ and $j*$ would give a stricter baseline for the causality. Thus, we conclude the $i*$ and $j*$ chosen are reasonable for the purpose in the current study.

Notes on linear trend

Prior to the model fit, the linear trend was removed to make the data segment quasi-stationary. However, the linear trend is useful to check if the %CC and NoTs anticorrelated during the task. We expected the %CC to decrease, while NoTs to increase with time, thus, a negative and positive linear trend for %CC and NoTs, respectively.

The linear trend was estimated applying a mixed-effect model, Y ~ Time (fixed variable) + participants (random variable) to $\%CC_t$ and $NoTs_t$ in of the low- and high %MW segment, separately. In the low %MW segment, the linear trend was negative for %CC; Linear coef.= -23.27, the 95%CI= [-32.20, -14.34], t= -4.97, and positive for NoTs; Linear coef.= 0.55, the 95%CI= [0.31, 0.78], t= 4.42. In the high %MW segment, the linear coefficient for $\%CC_t$was again negative; Linear coef.= -15.79, the 95%CI= [-23.66, -7.84], t= -3.80. For $NoTs_t$, the positive trend was not strong; Linear coef.= 0.18, the 95%CI= [-0.03, 0.39], t= 1.66. Overall, the results suggest that the %CC and NoTs anticorrelated during the task. The weak linear trend in the high %MW segment might suggests that the NoTs were near their ceiling; the

number of thoughts changed around 3.40 which is near 4, the limit in working memory capacity according to Luck and Vogel (1997). Moreover, this is above the 90% level of the standardized number of thoughts observed in our previous study (Nakatani et al., 2019). In contrast, the number varied around 2.72 in the low %MW segment.

Moreover, we checked the linear trend of AEP_t and $Alpha_t$. In the low %MW segment, the AEP amplitude decreased with time; Linear coef.= -0.28, the 95%CI= [-0.43, -0.14], t= -3.75, while the Alpha amplitude increased; Linear coef.= 0.42, the 95%CI= [0.21 0.64], t= 3.75. Similarly, in the high %MW segment, AEP amplitude decreased; Linear coef.= -0.37, the 95%CI= [-0.50, -0.24], t= -5.53, while Alpha amplitude increased with time; Linear coef.= 0.26, the 95%CI= [0.07, 0.46], t= 2.59.

The above results are in agreement with our assumption; the GDN and DMN activity overall have antagonistic effects to the cognitive control and spontaneous thought generation, but the antagonism is disrupted by the effect of $Alpha_{t-k}$, effective with a delay of 4-10 min (corresponding to *k=3,4 and 5*) on cognitive control.

Discussion

A good proportion of our thoughts does not have immediate relevance to our current task or goal (Antrobus, 1968; Giambra, 1995; Klinger & Cox, 1987). How do we switch to such a state of mind wandering from previously focused thought? The transition, when described as a dynamic trajectory, takes several intermediate stages (Nakatani et al., 2019), crossing into the territory of other spontaneous mental states (Christoff et al., 2016). We proposed that the trajectory is determined, at least to some extent, by oscillations in the balance between two antagonist neural systems: the Default Mode Network (DMN) and the goal-directed networks (GDN). We further hypothesized that the antagonism would be mitigated, in particular, during the transition from a goal-oriented/task-focused state to mind wandering. Breaks of the antagonism (and their eventual restoration) are, we believe, what makes the trajectory complex.

To explore how such breaks arise, in a tone counting task we firstly investigated if neural measures of DMN and GDN activity (indicated by Alpha and AEP amplitude, respectively) qualify as stochastically causal predictors of behavior, expressed as self-reported numbers of thoughts (NoTs), and correct tone counting performance (%CC), in data segments around the report of high %MW (mind wandering) and low %MW (task-focused) reports. Of the two neural measures, only the effect of Alpha was found stochastically causal. The effect of Alpha, as a measure of DMN activity, on %CC, our measure of cognitive control strength, changed from task-negative to positive between the low- and high %MW segments. This result was consistent with that in previous fMRI studies (See a review by Buckner & Dinicola, 2019). These suggested that the switch from task-negative to positive in DMN reflects a shift in mental focus from external task-related events to internal thoughts. The current EEG results could reflect the shift from external stimuli in the task-focused state to internal spontaneously generated thoughts. This observation would be in accordance with the first scenario sketched in the introduction, with respect to mitigating the antagonism of DMN and GDN.

Moreover, the current results suggested that Alpha activity several minutes before the high- or low %MW report had the strongest effect. The occurrence of effect latencies is in accordance with the

second mitigation scenario mentioned. Notably, the latencies were larger than observed in previous studies; In previous EEG studies, data segments of seconds before the subjective report of mind wandering or the response to the task were analyzed, e.g., the segments 0-12s before the report of mind wandering (Kirschner et al., 2012). fMRI studies have a larger time scale, because of the delay in the hemodynamic function (e.g., the default setting of the most widely used hemodynamic function in SPM is 6s); BOLD signal analyses were limited to within 1min from the reference point (e.g., Andrews-Hanna et al., 2014, a meta analysis). The current result reporting the relevance of past activity at the scale of several minutes therefore suggests that at least a part of DMN activity is involved in slow modulation of cognitive control.

In contrast, AEP showed no stochastic causal effect on any of our measures. It is generally assumed that AEP amplitude is related to the level of attention to the incoming stimulus (Picton & Hillyard, 1974; Picton et al., 1974). As a stimulus evoked potential, AEP could be less sensitive to sustained attention than ongoing activity (Dockree, Kelly, Foxe, Reilly, & Robertson, 2007). For example, Docree et al (2007) reported that visual evoked potential (VEP) amplitude was less sensitive to sustained attention than visual alpha-band amplitude. In the current dataset, however, the sensitivity between the AEP amplitude (full band) and the auditory alpha-band power did not differ (Bernhard, 2018). Thus, the current result may be explained better by the level of attention required by the tone counting task, which was low. Whereas level of attention is critical when processing is difficult, this is unlikely to be the case for the tone counting task. It was explicitly chosen to be easy and monotonous, in order to elicit mind wandering. At least under such conditions, the attentional gain offered by the GDN has no causal influence on the mental shift from task-focused to mind wandering states.

The term ‘Default-Mode Network’ originally referred to a group of brain regions with higher metabolic activity during rest than during task (Raichle, 2015; Raichle et al., 2001). Some of these regions initially had no distinctive functional attribution. Since then, as a result of functional mapping, an increasing number of subregions have been identified (Binder, Desai, Graves, & Conant, 2009; Buckner & Dinicola, 2019; Yeo et al., 2011). In this perspective, our collective label ‘DMN’ means little without considering the functions attributed to different sub-systems. An fMRI study reported contributions of various subsystems

to mind wandering (Kucyi, Esterman, Riley, & Valera, 2016). BOLD/fMRI signal from 7 subsystems (3 DMN subsystems: Core DMN, dorsomedial prefrontal cortex, and middle temporal lobule networks and 4 GDN ones: salience, dorsal attention, frontoparietal control, and sensorimotor networks) were correlated with the level of mind wandering (which participants rated in response to probes presented during a sustained visual attention task) and variability of response times (RT) to trials in the 30s intervals before the probes. In all DMN subsystems, the correlation between the BOLD signal and the level of mind wandering was positive, while that between the signal and the RT variance was negative (except for the medial temporal subsystem, which fell short of statistical significance). The latter result indicates that RTs were more stable when the DMN subsystems were highly activated. It, therefore, parallels with our observed task-positive DMN effect. Kucyi et al (2016) found no correlation between the BOLD signal of the GDN subsystems and the level of mind wandering. The authors attribute this to the unreliability of self-report. A positive correlation was found between the signal in the salience and dorsal attention networks and RT variability, which is in accordance with response variability studies (Esterman, Noonan, Rosenberg, & DeGutis, 2013). These results suggest that, as far as mind wandering is concerned, the respective subsystems of DMN operate in consort with each other, thereby validating the collective label over the subsystems.

By correlating neural data directly with subjective reports of mind wandering, the wider context of spontaneous mental process, in which mind wandering is situated (e.g., Christoff et al., 2016), would be lost. The two functions comprising Ψ, spontaneous thought generation and cognitive control, were uncorrelated to sufficient extent in our previous mind wandering study (Nakatani et al 2019) to be treated as the basis of a 2-dimensional state space, creating room for a variety of mental states.

Cognitive control in our experiment was modulated, not by GDN, but by DMN activity. This might appear to contradict the consensus in the literature, which links cognitive control to areas belonging to the GDN, such as the dorsal attention network, the salience network, and sensory-motor systems (Aben et al., 2017; Cohen et al., 1990; Funes et al., 2010; Ridderinkhof, 2002). However, there is no contradiction if we consider that the workload of the task was low and thus no strong contribution from GDN was required.

As for spontaneous thought generation, no stochastically causal neural effect was suggested. The only noticeable effect was that of $Alpha_t$ in the low %MW segment. This variable is instantaneous, and thus a correlate rather than a cause. The current results therefore provide little information about which neural activity modulates the rate of spontaneous thought generation.

Figure 3 illustrates implications of the current results. We hypothesized that effects of DMN and GDN activity on spontaneous thought generation and cognitive control are not fixed but variable, in particular, during the transition from task-focused to mind wandering states. The current results suggest that the DMN effect on cognitive control is stochastically causal and variable in sign. This could perturb the general anticorrelation between cognitive control and spontaneous thought generation (Figure 3, Right). The transient effect could appear as a deviation from the antidiagonal in the state space of spontaneous mental processes (Figure1 b).

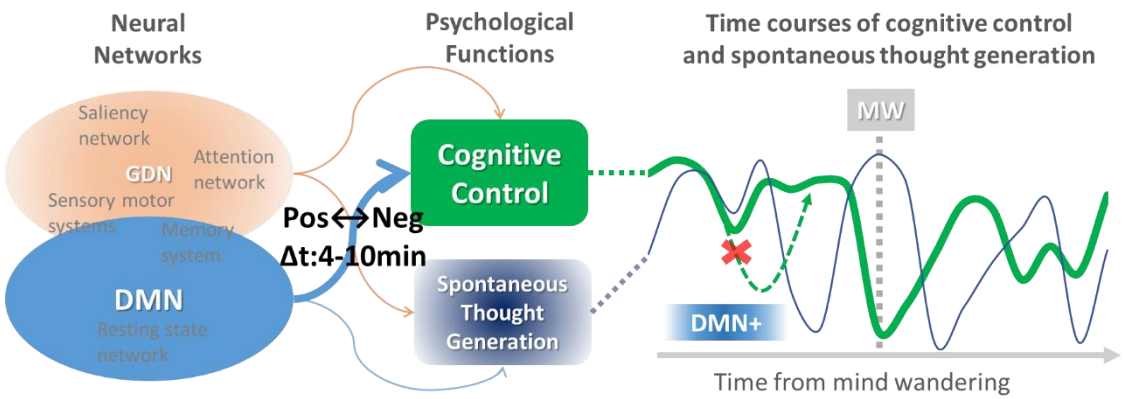


Figure 3. Estimated dynamic DMN effect (Left) and an example of the possible effect on the time course of the psychological function (Right)

The current results suggest that DMN effect on cognitive control is stochastically causal (Left, thick blue arrow). The direction of the effect varied between task positive and negative, depending on the context. These effects can lead to both anticorrelation and correlation between the cognitive control and spontaneous thought generation at different moments in time. E.g., the task-positivity (Right, 'DMN+') prior to the mind wandering report could prevent a dip in control (Right, green broken line). As a result, the control does not anticorrelate, but at least temporarily correlates positively with spontaneous thought generation.

The time scale of the current study agrees with our subjective experience of mind wandering. Nevertheless, investigation in this time scale has rarely been done. Situated in-between the time scales investigated in previous studies; seconds and shorter with neural measures (Braboszcz & Delorme, 2011; Ellamil et al., 2016; Kirschner et al., 2012; Kucyi et al., 2016), and hours and longer others with behavioral and subjective report measurements (Bennike, Wieghorst, & Kirk, 2017; Nakatani et al., 2019; Smith, Mills, Paxton, & Christoff, 2018), the current study offers a minimal framework for mind wandering and other spontaneous mental processes.

For the sake of monitoring fluctuations of mind wandering and task-focus for clinical or work safety purposes, or purely for the sake of ecological validity, it is of importance to study mind wandering outside of the laboratory. The current study offers a method that has potential for such applications. EEG recording using a mobile system suffices for measurement, as an EEG signal was able to detect neural effects in accordance with those reported in previous fMRI studies. With mobile EEG systems, the number of sensors is limited. The current study suggested that monitoring the right temporal alpha band activity could be sufficient to obtain a neural marker for mental states. The sensor location could differ depending on task modality. If necessary, the location can first be identified in the laboratory by applying the current data processing to other data with different task modality. Moreover, the main neural signal processing method was band-pass filtering which is relatively low in computational cost, and (quasi) real-time processing is possible. In addition to the sensor location and signal frequency, the current results specified the timing of possible marker signal, several minutes before the subjective report of %MW. Further analysis of the above signal, therefore, will lead to a reliable marker for the prediction and further understanding of cognitive-neuroscientific dynamics for mind wandering and related spontaneous mental states.

Open Practice Statement

All data are at the project page on Open Science Framework (https://osf.io/5ucvs/). They will be made publicly available after publication of the manuscript.

References

Aben, B., Verguts, T., & Van den Bussche, E. (2017). Beyond trial-by-trial adaptation: A quantification of the time scale of cognitive control. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 509-517. doi:10.1037/xhp0000324

Andrews-Hanna, J. R., Saxe, R., & Yarkoni, T. (2014). Contributions of episodic retrieval and mentalizing to autobiographical thought: Evidence from functional neuroimaging, resting-state connectivity, and fMRI meta-analyses. *Neuroimage, 91*, 324-335. doi:10.1016/j.neuroimage.2014.01.032

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the new York Academy of Sciences, 1316*(1), 29-52.

Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *British Journal of Psychology, 59*(4), 423-430.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Bennike, I. H., Wieghorst, A., & Kirk, U. (2017). Online-based Mindfulness Training Reduces Behavioral Markers of Mind Wandering. *Journal of Cognitive Enhancement, 1*(2), 172-181. doi:10.1007/s41465-017-0020-9

Bernhard, H. (2018). *Tracking the mind: measuring the cortical dynamics during mind wandering in an auditory counting task. .* (Research Master). Maastricht University, Maastricht, The Netherlands.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex, 19*(12), 2767-2796.

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage, 54*(4), 3040-3047.

Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., . . . Nagarajan, S. S. (2011). Measuring functional connectivity using MEG: methodology and comparison with fcMRI. *Neuroimage, 56*(3), 1082-1104.

Buckner, R. L., & Dinicola, L. M. (2019). The brain's default network: updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience, 20*(10), 593-608. doi:10.1038/s41583-019-0212-7

Chen, A. C., Feng, W., Zhao, H., Yin, Y., & Wang, P. (2008). EEG default mode network in the human brain: spectral regional field powers. *Neuroimage, 41*(2), 561-574.

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

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience, 17*(11), 718-731.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332-361. doi:10.1037/0033-295x.97.3.332

Cole, M. W., & Schneider, W. (2007). The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage, 37*(1), 343-360.

Compton, R. J., Gearinger, D., & Wild, H. (2019). The wandering mind oscillates: EEG alpha power is enhanced during moments of mind-wandering. *Cognitive, Affective, & Behavioral Neuroscience, 19*(5), 1184-1191. doi:10.3758/s13415-019-00745-9

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience, 3*(3), 292-297. doi:10.1038/73009

Dockree, P. M., Kelly, S. P., Foxe, J. J., Reilly, R. B., & Robertson, I. H. (2007). Optimal sustained attention is linked to the spectral content of background EEG activity: Greater ongoing tonic alpha (~ 10 Hz) power supports successful phasic goal activation. *European Journal of Neuroscience, 25*(3), 900-907.

Ellamil, M., Fox, K. C., Dixon, M. L., Pritchard, S., Todd, R. M., Thompson, E., & Christoff, K. (2016). Dynamics of neural recruitment surrounding the spontaneous arising of thoughts in experienced mindfulness practitioners. *Neuroimage, 136*, 186-196. doi:10.1016/j.neuroimage.2016.04.034

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

Fink, A., Graif, B., & Neubauer, A. C. (2009). Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers. *Neuroimage, 46*(3), 854-862. doi:10.1016/j.neuroimage.2009.02.036

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

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the resting-state default mode of brain function hypothesis. *Human Brain Mapping, 26*(1), 15-29. doi:10.1002/hbm.20113

Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Sustained vs. Transient cognitive control: Evidence of a behavioral dissociation. *Cognition, 114*(3), 338-347. doi:10.1016/j.cognition.2009.10.007

Giambra, L. M. (1995). A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness and cognition, 4*(1), 1-21.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424-438. doi:10.2307/1912791

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences, 100*(1), 253-258. doi:10.1073/pnas.0135058100

Gruberger, M., Simon, E., Levkovitz, Y., Zangen, A., & Hendler, T. (2011). Towards a Neuroscience of Mind-Wandering. *Frontiers in human neuroscience, 5*(56). doi:10.3389/fnhum.2011.00056

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*(3), 284-291. doi:10.1038/72999

Jaarsveld, S., & Van Leeuwen, C. (2005). Sketches from a Design Process: Creative Cognition Inferred From Intermediate Products. *Cognitive science, 29*(1), 79-101. doi:10.1207/s15516709cog2901_4

Jerbi, K., Vidal, J. R., Ossandon, T., Dalal, S. S., Jung, J., Hoffmann, D., . . . Lachaux, J.-P. (2010). Exploring the electrophysiological correlates of the default-mode network with intracerebral EEG. *Frontiers in Systems Neuroscience*. doi:10.3389/fnsys.2010.00027

Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2019). Predicting task-general mind-wandering with EEG. *Cognitive, Affective, & Behavioral Neuroscience, 19*(4), 1059-1073. doi:10.3758/s13415-019-00707-1

Kirschner, A., Kam, J. W. Y., Handy, T. C., & Ward, L. M. (2012). Differential synchronization in default and task-specific networks of the human brain. *Frontiers in human neuroscience, 6*, 139.

Klinger, E., & Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagination, Cognition and Personality, 7*(2), 105-128.

Kucyi, A., Esterman, M., Riley, C. S., & Valera, E. M. (2016). Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences, 113*(48), 13899-13904. doi:10.1073/pnas.1611743113

Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., & Kleinschmidt, A. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proceedings of the National Academy of Sciences, 100*(19), 11053-11058. doi:10.1073/pnas.1831638100

Lehtelä, L., Salmelin, R., & Hari, R. (1997). Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. *Neuroscience Letters, 222*(2), 111-114. doi:10.1016/s0304-3940(97)13361-4

Luca, C., Andrew, Z., Alex, F., & Jason, B. M. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in Cognitive Sciences, 17*(10), 493-501. doi:https://doi.org/10.1016/j.tics.2013.08.006

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281. doi:10.1038/36846

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., & Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences, 104*(32), 13170-13175. doi:10.1073/pnas.0700668104

Marino, M., Arcara, G., Porcaro, C., & Mantini, D. (2019). Hemodynamic Correlates of Electrophysiological Activity in the Default Mode Network. *Frontiers in Neuroscience, 13*(1060). doi:10.3389/fnins.2019.01060

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science, 315*(5810), 393-395.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function, 214*(5-6), 655-667.

Nakatani, C., Ganschow, B., & van Leeuwen, C. (2019). Long-term dynamics of mind wandering: ultradian rhythms in thought generation. *Neuroscience of Consciousness, 2019*(1). doi:10.1093/nc/niz007

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive Control Predicts Use of Model-based Reinforcement Learning. *Journal of cognitive neuroscience, 27*(2), 319-333. doi:10.1162/jocn_a_00709

Pearl, J. (2000). Causality: Models, reasoning and inference cambridge university press. *Cambridge, MA, USA, 9*, 10-11.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and clinical neurophysiology, 36*, 191-200.

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and clinical neurophysiology, 36*, 179-190.

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience, 38*(1), 433-447. doi:10.1146/annurev-neuro-071013-014030

Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676-682. doi:10.1073/pnas.98.2.676

Ridderinkhof, R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research, 66*(4), 312-323. doi:10.1007/s00426-002-0104-7

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience, 27*(9), 2349-2356. doi:10.1523/jneurosci.5587-06.2007

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*(6), 946-958. doi:10.1037/0033-2909.132.6.946

Smith, G. K., Mills, C., Paxton, A., & Christoff, K. (2018). Mind-wandering rates fluctuate across the day: evidence from an experience-sampling study. *Cogn Res Princ Implic, 3*(1), 54. doi:10.1186/s41235-018-0141-4

Tiihonen, J., Hari, R., Kajola, M., Karhu, J., Ahlfors, S., & Tissari, S. (1991). Magnetoencephalographic 10-Hz rhythm from the human auditory cortex. *Neuroscience Letters, 129*(2), 303-305. doi:10.1016/0304-3940(91)90486-d

Voss, U., & Voss, G. (2014). A neurobiological model of lucid dreaming. In R. Hurd & K. Bulkeley (Eds.), *Lucid Dreaming: New Perspectives on Consciousness in Sleep [2 volumes]: New Perspectives on Consciousness in Sleep* (Vol. 1, pp. 23-36). Santa Barbara, CA: ABC-CLIO.

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., & Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Front Psychol, 2*, 73. doi:10.3389/fpsyg.2011.00073

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., . . . Polimeni, J. R. (2011). The Organization of the Human Cerebral Cortex Estimated by Functional Connectivity 2. *network, 21*, 22.

Appendix

1. Mind wandering percentage (%MW) histogram

Figure A1 provides the %MW histogram over all participants. Low-, intermediate-, and high %MW trials are marked in yellow (%MW<30), gray (40≦%MW<60), and purple (%MW≧70), respectively

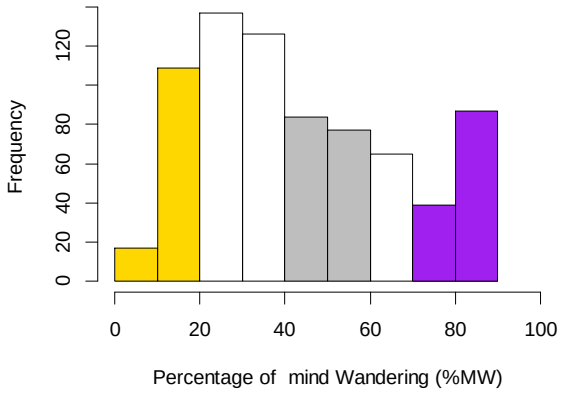


Fig A1. Histogram of the reported mind wandering percentage (%MW)

2. Electrodes for auditory evoked potential (AEP) and Alpha amplitudes

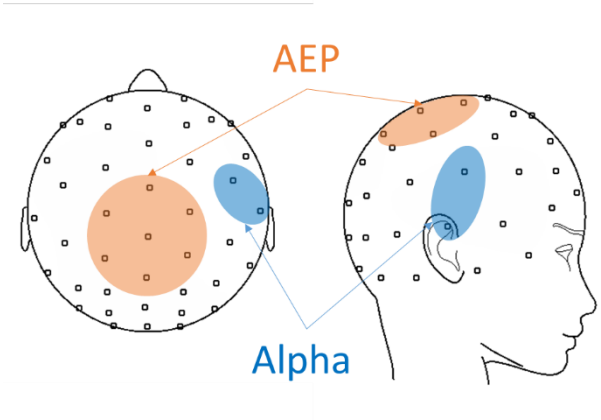


Figure A2. Electrodes for auditory evoked potential (AEP) and Alpha amplitudes
Electrodes used to compute the AEP and Alpha amplitude are indicated. Note that the electrode placement is not the international 10-10, but a manufacturer-defined system.

3. Auto regression effects at r=1, s=1

The effect of autoregression at *i=1, j=1* was estimated to compare that with pre-selected *i** and *l** (See main text Notes on autoregression effect). In the same way as that in models for Tables 2, 3, and 5 in main text, the index for AEP was set as *l=1*.

a) Estimated coefficients of the full model with (*i=1, j=1, k=4, l=1*) for high %MW segments

model: $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-1}) + (Alpha_{t-4} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC$_t$**	(Intercept)	-4.39	-7.07	-1.70	-3.21	0.0014
	NoTs $_{t-1}$	2.02	-1.48	5.52	1.14	0.2571
	%CC $_{t-1}$	0.18	0.09	0.27	3.88	0.0001
	Alpha $_{t-4}$	8.07	4.11	12.04	4.00	0.0001
	AEP $_{t-1}$	2.46	-3.43	8.35	0.82	0.4124
	Alpha $_t$	-0.72	-4.45	3.00	-0.38	0.7031
	AEP $_t$	3.90	-2.02	9.83	1.30	0.1960
	Participants	0.36	0.14	0.59	3.16	0.0017

Response var. = **NoTs $_t$**	(Intercept)	0.00	-0.07	0.07	-0.11	0.9120
	NoTs $_{t-1}$	0.19	0.10	0.28	4.06	0.0001
	%CC $_{t-1}$	0.00	0.00	0.01	2.44	0.0150
	Alpha $_{t-4}$	0.07	-0.04	0.17	1.28	0.2020
	AEP $_{t-1}$	-0.12	-0.27	0.03	-1.54	0.1240
	Alpha $_t$	0.08	-0.02	0.18	1.64	0.1030
	AEP $_t$	0.11	-0.04	0.26	1.41	0.1600
	Participants	0.00	0.00	0.01	0.98	0.3280

b) Estimated coefficients of the full model with (*i=1, j=1, k=5, l=1*) for high %MW segments

model: $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-1}) + (Alpha_{t-5} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC$_t$**	(Intercept)	-4.50	-7.20	-1.80	-3.28	0.0011
	NoTs $_{t-1}$	1.75	-1.78	5.29	0.98	0.3293
	%CC $_{t-1}$	0.17	0.08	0.26	3.56	0.0004
	Alpha $_{t-5}$	6.74	2.75	10.72	3.32	0.0010
	AEP $_{t-1}$	3.82	-2.09	9.74	1.27	0.2047
	Alpha $_t$	0.19	-3.50	3.87	0.10	0.9212
	AEP $_t$	2.68	-3.27	8.63	0.88	0.3769
	Participants	0.38	0.15	0.60	3.28	0.0011

Response var. = **NoTs $_t$**	(Intercept)	0.00	-0.07	0.06	-0.13	0.8939
	NoTs $_{t-1}$	0.18	0.09	0.28	3.98	0.0001
	%CC $_{t-1}$	0.00	0.00	0.01	2.30	0.0217
	Alpha $_{t-5}$	0.06	-0.04	0.16	1.19	0.2362
	AEP $_{t-1}$	-0.11	-0.26	0.04	-1.39	0.1656
	Alpha $_t$	0.09	-0.01	0.18	1.80	0.0718
	AEP $_t$	0.10	-0.05	0.25	1.28	0.2030
	Participants	0.00	0.00	0.01	1.02	0.3088

c) Estimated coefficients of the full model with (*i=1, j=1, k=3, l=1*) for Low %MW segments

model: $[NoTs_t, \%CC_t] \sim (NoTs_{t-1} + \%CC_{t-1}) + (Alpha_{t-3} + AEP_{t-1}) + (Alpha_t + AEP_t) + Participants$

	Predictors	Coefficients	Confidence interval 2.50%	97.50%	t-value	Pr(>|t|)
Response var. = **%CC$_t$**	(Intercept)	-6.09	-9.89	-2.30	-3.16	0.0017
	NoTs $_{t-1}$	6.46	2.36	10.55	3.10	0.0021
	%CC $_{t-1}$	0.11	0.01	0.21	2.17	0.0310
	Alpha $_{t-3}$	-10.48	-15.12	-5.84	-4.44	0.0000
	AEP $_{t-1}$	1.53	-4.98	8.04	0.46	0.6449
	Alpha $_t$	7.30	2.64	11.96	3.08	0.0022
	AEP $_t$	-2.23	-8.89	4.43	-0.66	0.5110
	Participants	0.52	0.21	0.83	3.26	0.0012

Response var. = **NoTs $_t$**	(Intercept)	-0.05	-0.14	0.03	-1.21	0.2285
	NoTs $_{t-1}$	0.14	0.04	0.23	2.88	0.0042
	%CC $_{t-1}$	0.00	0.00	0.00	2.02	0.0440
	Alpha $_{t-3}$	0.10	0.00	0.21	1.92	0.0558
	AEP $_{t-1}$	0.02	-0.13	0.17	0.29	0.7758
	Alpha $_t$	-0.17	-0.27	-0.06	-3.13	0.0019
	AEP $_t$	0.03	-0.12	0.18	0.37	0.7103
	Participants	0.01	0.00	0.01	1.75	0.0804