**Intentionality of Mind-Wandering as Reflected in Measures of Executive Control and Behavioral Variability: a TMS Study**

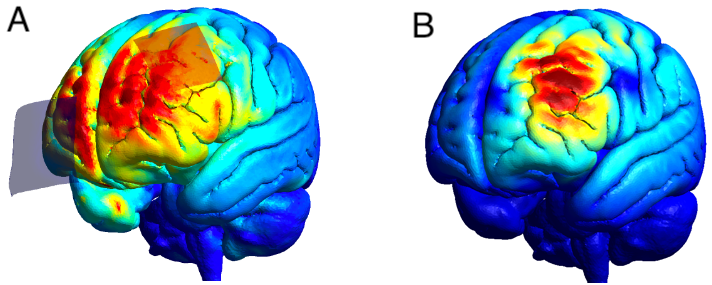
1. **Introduction**

Humans spend a substantial amount of their waking lives engaged in spontaneous, self-generated thoughts that are decoupled from an ongoing activity or the current surroundings (Killingsworth & Gilbert, 2010; Seli et al., 2018). This mental phenomenon has been studied under the umbrella term of “mind-wandering” (MW) (Kane et al., 2007; Killingsworth & Gilbert, 2010; Klinger & Cox, 1987). Over the past two decades, cognitive neuroscientists have increasingly gained interest in elucidating the basic neurocognitive mechanisms and physiological underpinnings of MW (Callard et al., 2013). Interestingly, whilst MW has been associated with future planning and creative problem-solving (Mooneyham & Schooler, 2013), it has also been shown to interfere with task performance (Smallwood & Schooler, 2015) and negatively impact emotional well-being (Hoffmann et al., 2016).

It has been argued that MW, when detrimental to task performance, is the result of disruption of executive control (McVay & Kane, 2010; Smallwood & Schooler, 2006). However, the task which has long dominated MW research, the sustained-attention-to-response-task (SART: (Robertson et al., 1997), is not fit to measure executive control since it was designed to test only one of constituent processes thereof: response inhibition. However, executive control also involves constant monitoring of task progress which recruits working memory. Hence, this study harnessed the finger-tapping random sequence generation task (FT-RSGT), recently designed and validated by (Nya Mehnwolo Boayue et al., 2019).

The central role of executive control in MW has also been revealed by neuroimaging studies: the recruitment of executive and default mode networks during MW is reflected in significant activation patterns within their key prefrontal hubs (Christoff et al., 2009; Turnbull et al., 2019). These accounts have prompted another line of work closely related to the present study: neuromodulation of MW by means of non-invasive brain stimulation (NIBS). To this day, several studies have attempted to influence the propensity to mind-wander by means of transcranial direct current stimulation (tDCS). The yielded results are inconclusive: whilst some studies reported that anodal (excitatory) tDCS over left dorso-lateral prefrontal cortex (dlPFC) induced an increase in MW (Axelrod et al., 2015, 2018), others reported the same effect for cathodal (inhibitory) tDCS over the same region (Filmer et al., 2019), yet others showed no effect of anodal tDCS on MW likely caused by the weak modulatory effect of brain polarization procedures modulating regional excitability rather than operating on brain rhythms (Boayue et al., 2019).

This evidence taken together points at subtle methodological inconsistencies. Firstly, the targeted region varied greatly across studies. Secondly, tDCS, while easy to implement, is characterized by insufficient focality: as demonstrated by a SimNIBS simulation of the montage in (Axelrod et al., 2015), the current propagates far beyond the target region, left dlPFC (fig. 1).



**Figure 1.** **A**: SimNIBS simulation of the montage tested in Axelrod et al. (2015). The electric field propagates beyond the ROI (left dlPFC). **B**: SimNIBS simulation of the TMS protocol of the present study (MSO 55%). The impact of the magnetic field remains within the bounds of the dorso-lateral regions of the left hemisphere.

The objective of this study was to elucidate the relationship between non-invasive brain stimulation (NIBS) and MW by testing the ability of transcranial magnetic stimulation (TMS) to modulate MW states. More specifically, we investigated the impact of theta entrainment via rhythmic TMS on MW propensity and task measures. The emphasis on theta entrainment as the modulator of MW stems from the literature identifying fronto-medial theta oscillations as the underlying marker of sustained attention (Clayton et al., 2015) and cognitive control (Cavanagh & Frank, 2014). We attempted to answer the following questions: can TMS over left dlPFC modulate MW propensity? By extension, can it also have an impact on task performance and executive control as reflected in behavioral measures? To our knowledge, this study is the first in MW research to use online TMS and to probe for a causal impact of oscillatory theta entrainment on task performance. To account for placebo effect and to isolate the effect of entrainment versus mere modulation of excitability, we employed 4 control conditions: sham rhythmic TMS (rhTMS), active arrhythmic TMS (arrhTMS), sham arrhythmic TMS and baseline, devoid of any stimulation. The outcome measures, experimental design and operational definitions are outlined hereunder. This study was preregistered on OSF platform: <https://osf.io/2wszr>.

We tested the following hypotheses:

* H1. Based on the correlation of fronto-medial theta oscillations with sustained attention (Clayton et al., 2015) and cognitive control (Cavanagh & Frank, 2014), we expect participants to mind-wander less when subjected to active rhTMS compared to active arrhTMS, sham stimulation and baseline. Quantitatively, we expected a positive effect of active rhythmic TMS on task scores (a higher score corresponds to less MW) as reflected in the coefficient value within the fitted model.
* H2. We hypothesize an increase in executive control during active rhythmic TMS compared to other conditions. Quantitatively, this would manifest in an increase in approximate entropy (AE).
* H3. By extension, we expect subjects to more accurately emulate the rhythm of the metronome with finger taps during active rhythmic TMS compared to other conditions. This would reflect in lower behavioral variability (BV).

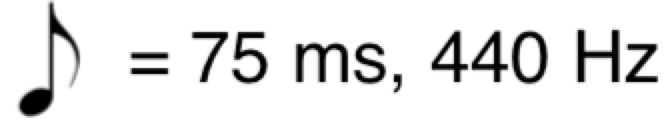
Considering the relative novelty of the FT-RSGT, prior to the main TMS-EEG experiment, we first attempted to replicate the key features of the task as previously done in Boayue et al. (2021) and to thus validate it. To this end, we conducted a behavioral pilot whereby 13 participants completed 6 blocks of the task, 10 minutes each. Since the pilot was not part of the main study, we report the results as exploratory analyses the aim of which was to further lend support to content validity of the task and to demonstrate that performance metrics of FT-RSGT have indeed the potential to reflect MW states.

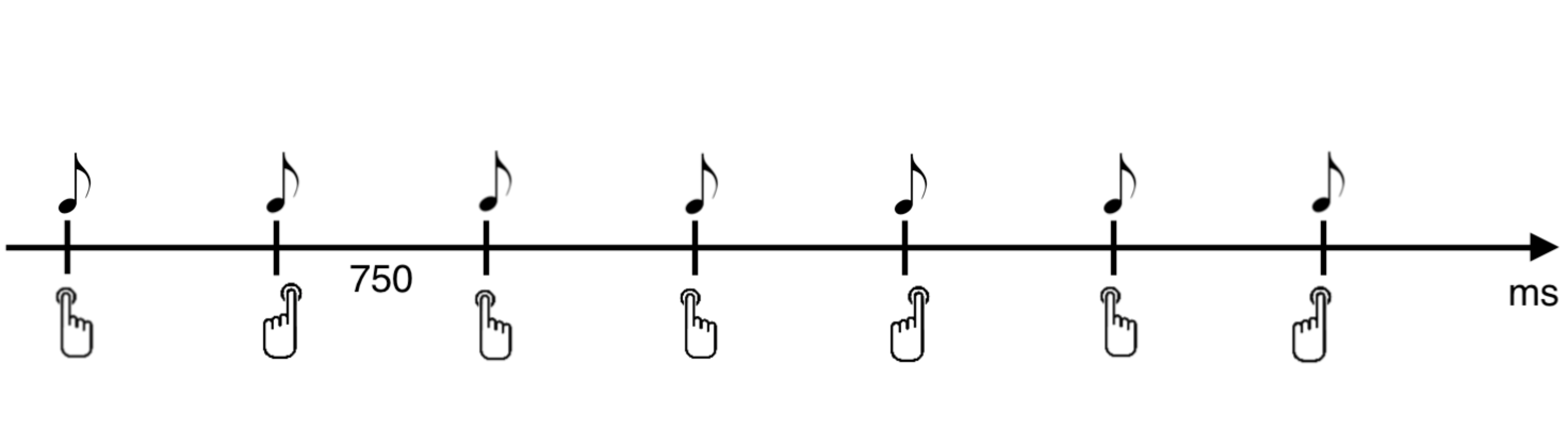
1. **Methods**
   1. **Behavioral Pilot**
      1. ***Participants***

The pilot sample consisted of 11 members (6 females) of the Frontlab laboratory and employees of the Paris Brain Institute aged 21 to 32 years old (mean age = 24.75, SD = 3.25). The only inclusion criterion was physical ability to complete the task and normal or corrected to normal vision.

* + 1. ***Behavioral task & Procedure***

This study involves a novel task – the Finger-Tapping Random Sequence Generation Task (FT-RSGT: Boayue et al., 2021). FT-RSGT is a combination of a modified version of the random generation task (Baddeley et al., 1998; Towse, 1998) and a finger-tapping task (Kucyi et al., 2016; Seli et al., 2013). This task consists of two sub-tasks: i. rhythmic finger-tapping in response to an ongoing metronome and ii. the generation of irregular sequences by pressing one of the two available response-buttons (“S”, left key and “L”, right key). Subjects were instructed to emulate the rhythm of the metronome as accurately as possible with their finger taps and, simultaneously, to try and render every button press as unpredictable for the external observer as possible. To ensure that the participants understand the task, we provided ample examples of how an irregular sequence would compare to a regular sequence (e.g. “right-left-right-left” is more regular than “right-right-left-right”) and emphasized that each press must be difficult to predict for an external party. The subjects also underwent a training upon which they were asked to retrospectively assess the tapping sequences they had produced and to provide examples thereof. Figure 2 represents a schematic of the task.

******

****

**Figure 2.** FT-RSGT: participants are instructed to press the right (L) or left (S) key in an irregular order simultaneously with the rhythm of the metronome. The tone of the metronome has a frequency of 440 Hz and its duration is 75 ms. ISI = 750 ms.

Given that the generation of irregular sequences draws heavily on executive resources, the randomness of the generated sequence is related to their deployment. This has been confirmed by the finding that sequences generated during MW are typically less random (Boayue et al., 2020; Teasdale et al., 1995). In addition, the behavioral variability as measured by deviation of the taps from the on-going metronome in the finger-tapping studies have been shown to be an indicator of MW (Kucyi et al., 2016; Seli et al., 2013).

During the experiment, the participants were seated in a chair, 57 centimeters away from the screen. The room was dimly lit, and participants wore a headset. Each trial began with a tone of 440 Hz lasting for 75 ms as indicated in figure 2. The tone repeated every 750 ms (= inter-stimulus interval; ISI) until the appearance of a thought-probe (see the section on measures). The ISI of 750 ms was validated by (Boayue et al., 2019) whereby they demonstrated that this interval was long enough for the executive control to be deployed, but also short enough so that the attention is maintained. Participants underwent 6 blocks of the task, 10 minutes each. The task was completed either in French or in English.

***2.1.3 Measures***

*2.1.3.1 Mind-Wandering*

During each the task, subjects were probed randomly by task-embedded experience sampling – also known as thought-probes. Quantitatively, it was assessed via task-embedded experience sampling: during the task, participants were interrupted by thought-probes which appeared every 40 to 80 seconds. Thought-probes asked participants to evaluate their MW propensity (fig. 3). Over the course of the experiment, 60 thought-probes were answered which yielded 60 measures for each variable of interest for each subject.

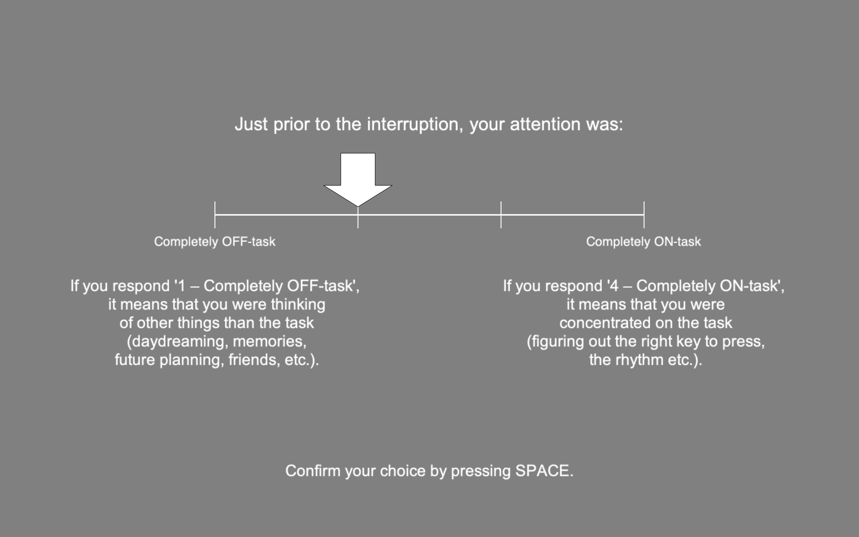


Figure 3

**Figure 3.** The thought-probe was presented to participants randomly throughout the task (every 40 to 80 minutes). The initial position of the arrow was randomized so as to avoid participant priming. Participants used keys “S”and “L” to move the arrow along the scale and confirmed their answer by pressing “space”.

The questions were accompanied by instructions providing specific examples of mental states to which the ratings would pertain (figure 2). The answer yielding a score of 1 would indicate maximum MW or a completely off-task state.

*2.1.3.2 Behavioral Variability*

Behavioral variability (BV) is used in this study as a measure reflecting task performance. BV is measured as the deviation of the rhythm of button presses from the metronome. BV has been previously associated with increased MW propensity (Boayue et al., 2021; Kucyi et al., 2016; Seli et al., 2013).

*2.1.3.3 Approximate Entropy*

Approximate Entropy (AE: Pincus, 1991) is a measure of randomness of a finite sequence of elements and, in this study, represents executive control. We used AE as the measure of the irregularity of the sequence of left-right presses produced by participants. Mathematically, AE(*i*) is a function of the number of elements (*i*) in the sequence. It yields the frequency with which blocks of length *i* remain close to each other. This frequency reflects the predictability of the sequence which changes with time as more elements are fed into the sequence. From a practical standpoint, AE is an inference of the predictability of the *i*th item in the sequence based on the predictability of [*i – 1*] items.

***2.2. Online rhythmic TMS-EEG Experiment***

*2.2.1. Participants*

We conducted an a-priori power analysis for a repeated-measure, within-factor ANOVA using G\*Power (also implemented in Gouraud et al., (2018). To reach the minimum acceptable power of 0.8 and to detect a medium effect (f = 0.25), a sample of 21 participants would be required (N of measures = 5). However, due to feasibility concerns and time constraints, we only managed to recruit 10 subjects. We prioritized those subjects whose T1 MRI scans had already been acquired. Our target population is in good health, right-handed, aged between 18 and 65 years old, fluent in written French or English and eligible according to MRI and TMS international safety guidelines.

However, we did not include people to whom at least one of the following pertains:

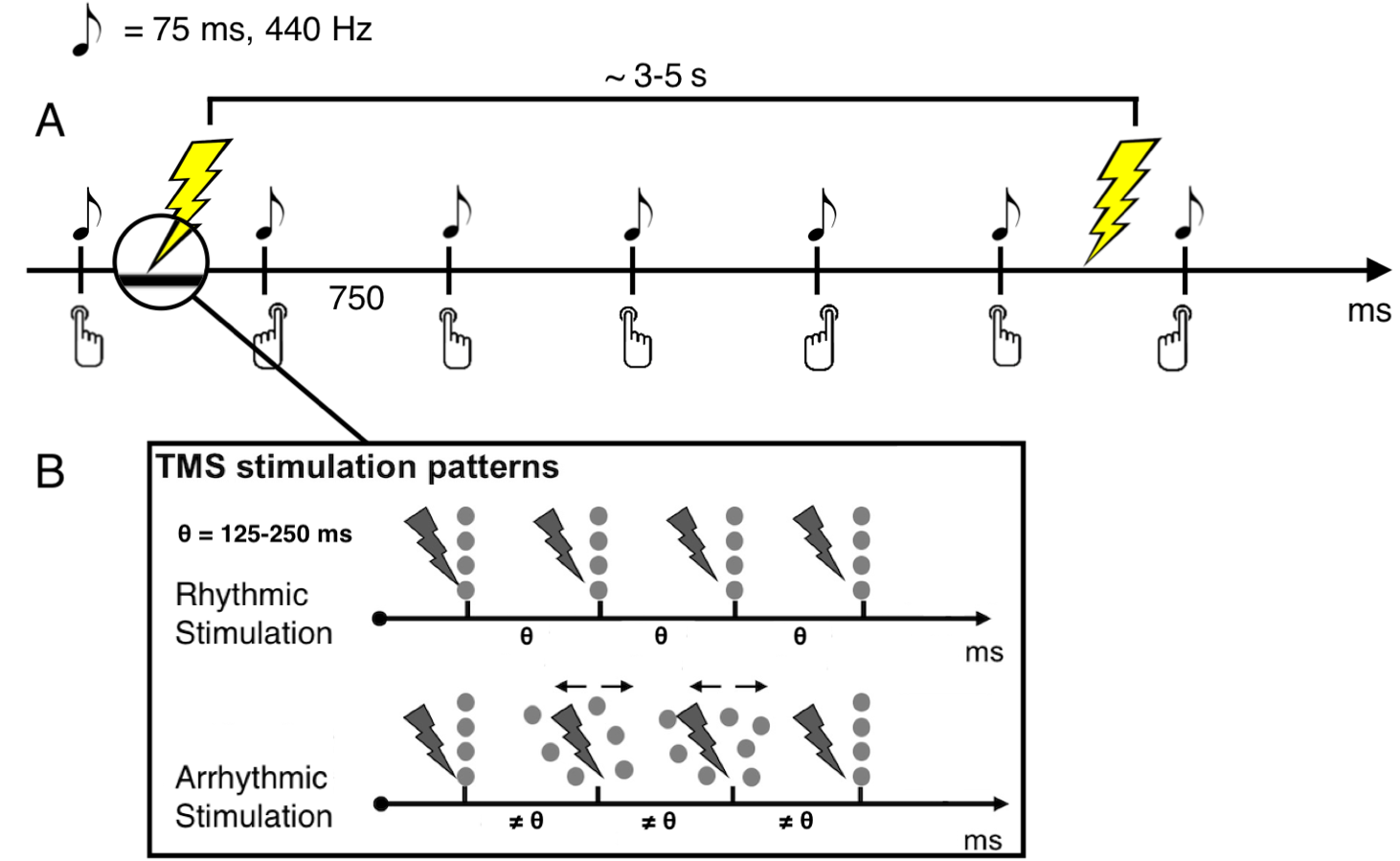
* currently participating in another study (< 24 hours or 1 week for studies involving brain stimulation or any other intervention affecting brain excitability);
* presenting or having a history of a psychiatric or neurological disorder or evolutive disease that interferes with the study tests;
* reported consumption of psychotropic substances (except nicotine and caffeine);
* taking central nervous medications (e.g. antidepressants, antiepileptic drugs) under benzodiazepines, anticonvulsants or neuroleptics treatment;
* pregnant, breastfeeding or has recently given birth;
* presenting a contra-indication to MRI.

We further excluded people who ask to stop the experiment or fail to cooperate and/or comply with the procedures during the experiment. The data of 8 subjects were analyzed.

*2.2. Experimental Procedure*

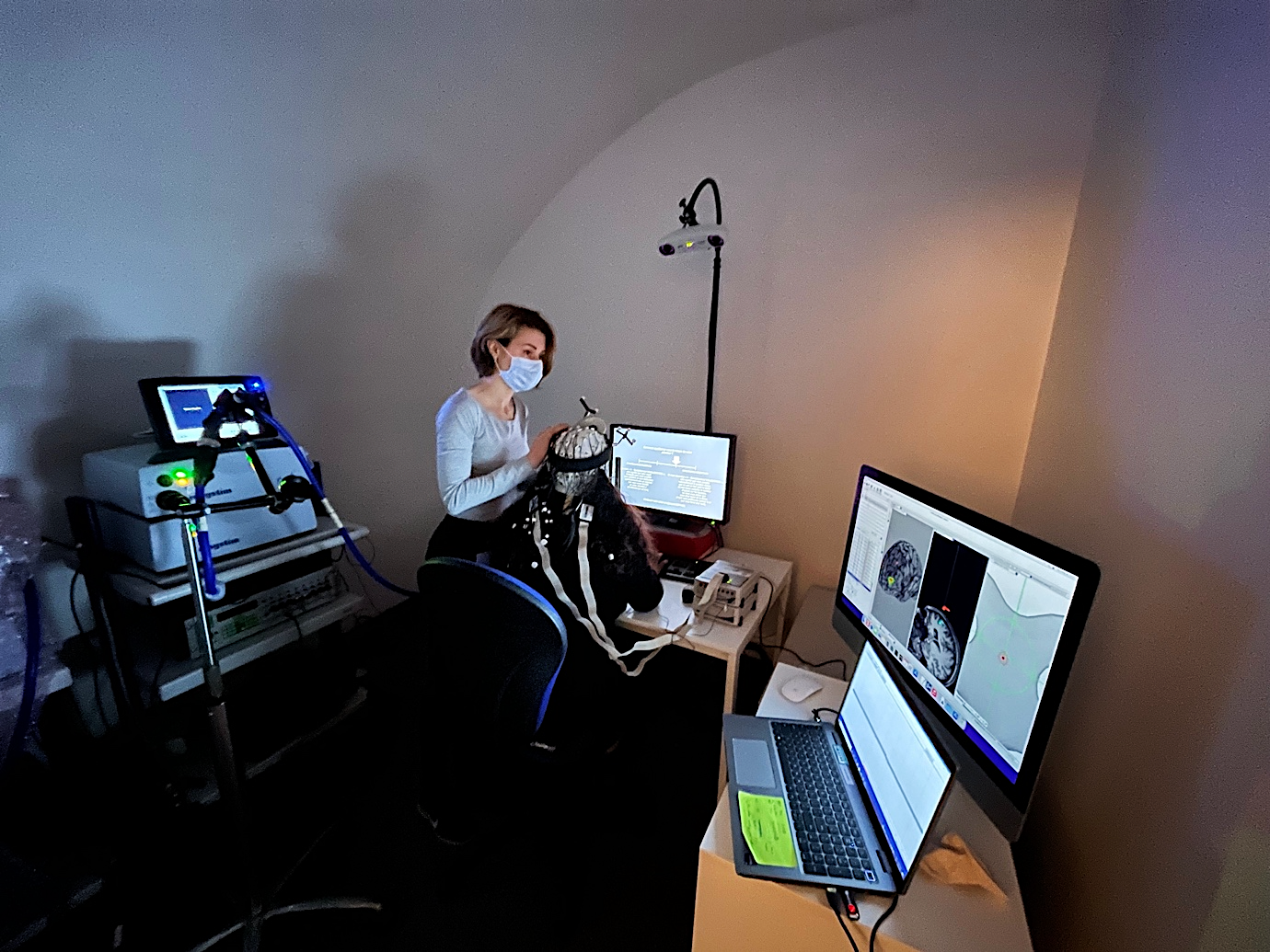
Participants were seated in a comfortable chair, with their head resting on a chinrest at a distance of 57 cm from the screen. The task script ran on PsychoPy. The same script was used to trigger TMS pulses. Both TMS triggering and EEG event marking was done via Arduino UNO. The task began with instructions during which participants were encouraged to ask questions if anything was unclear. Participants were instructed to place their index fingers on two keyboard keys (“S” and “L”) and to fixate the cross in the centre of the screen throughout the entire experiment. Henceforth, the task didn’t differ from the one used for the behavioral pilot (section 2.1.2.).

The schematic representation of the task along with TMS bursts is depicted in figure 4. The experiment consisted in two visits whereby the subject completed 14 blocks (= 110 mins) of FT-RSGT in total. The following section provides a detailed overview of the block design.



**Figure 4.** **A**: FT-RSGT: participants are instructed to press the right (L) or left (S) key in an irregular order simultaneously with the rhythm of the metronome. The tone of the metronome has a frequency of 440 Hz and its duration is 75 ms. TMS is administered every 3 to 5 s. **B**: Every TMS burst consisted of four pulses. Two stimulation patterns were implemented: rhythmic (rhTMS, top) and arrhythmic (arrhTMS, bottom). The inter-pulse interval (IPI: θ) was set based on the frequency of pulses. The latter was determined for each subject separately: an individual theta-peak frequency was extracted from the EEG recording of the first baseline. As a result, the IPI fell in the range of 125-250 ms (4-8 Hz). In the case of the arrhTMS, the IPI ≠ θ, but the total duration of the burst (3 \* θ) was identical for both patterns. However, IPI was always greater than 20 ms since it is the minimum IPI for Magstim Rapid.

EEG data were also collected during the experiment. However, they were not analyzed in the context of this study. The experimental setup is depicted in figure 5.



**Figure 5.** Experimental setup. The subject was wearing a 64-electrode EEG cap with active electrodes (actiCHamp Plus, Brain Vision LLC). Precise TMS targeting was ensured via Brainsight TMS Navigation system: an infrared camera tracked the position of the coil and the subject’s head, both equipped with trackers. During active TMS blocks the experimenter was holding the coil and adjusting its position via the coil-centered tracking view on the right.

***2.4. Stimulation Protocol***

The participants were subjected to 5 conditions in total over the course of two visits: baseline, sham rhythmic TMS (rhTMS), sham arrhythmic TMS (arrhTMS), active rhTMS and active arrhTMS. The condition of interest being active rhTMS, the other conditions serve as controls: while sham stimulation controls for the side effects of active stimulation only, the arrhythmic TMS allows to control for the potential effect of the frequency of the entrained oscillation and keep the side-effects accompanying active TMS (possible muscle twitching, noise etc.). The arrhythmic control condition thus allows to isolate the variable of interest (theta oscillations) and to preserve participant blinding. It has been also argued that sham stimulation on its own lacks specificity to be regarded as a full-fledged control condition (Duecker & Sack, 2015).

The subjects were sequentially randomized and assigned to either the rhythmic or arrhythmic group. During the first visit, the rhythmic group first underwent rhTMS followed by arrhTMS. For the arrhythmic group this order was reversed. The order of TMS conditions was always reversed for the second visit. Thus, each participant was exposed to all five conditions at both visits. The experimental protocol is outlined in table 1.

**Table 1.** Experimental protocol of the study. The subjects underwent two sessions of FT-RSGT and online TMS. The order of rhythmic and arrhythmic conditions was reversed for the visits. The subjects in the rhythmic group were administered rhythmic TMS during blocks 2 and 3 and arrhythmic TMS during blocks 5 and 6 at the first visit. In the arrhythmic group and for the second visit the order was reversed. In total, the protocol yielded 150 mins of task time (12000 trials).

|  |  |  |
| --- | --- | --- |
| **Block** | **Visit 1 or 2 ≈ 2h** | **Duration** |
| 1 | Baseline | 5’ |
| 2 | Sham rh/arrhTMS | 15’ |
| 3 | Active rh/arrhTMS | 15’ |
| 4 | Baseline | 5’ |
| 5 | Sham arrh/rhTMS | 15’ |
| 6 | Active arrh/rhTMS | 15’ |
| 7 | Baseline | 5’ |

TMS was applied with a biphasic TMS device (Magstim Super Rapid) equipped with a figure-of-eight coil (Double 70-mm Alpha Coil; The Magstim Company Ltd, UK). All TMS applications followed the updated safety guidelines and recommendations of the international TMS community (Rossi et al., 2021).

*2.4.1. MRI Preprocessing & ROI*

The ROI marking in the native space was done with SPM (Matlab). Each T1 scan was resliced and normalized. Left dlPFC was marked onto a normalized (MNI152) version of the scan. The corresponding VOI was then de-normalized and co-registered with the resliced T1 scan. We used the following MNI coordinates to define the ROI: x = -37, y = 41, z = 22 (Groot et al., 2020).

We used a frameless stereotactic system (*Brainsight TMS Navigation*) to minimize any deviations of the coil from the targeted site. The ROI was marked with SPM (Matlab) on the subject’s T1 scan in the native space based on the MNI coordinates derived by (Groot et al., 2020): x = -37, y = 41, z = 22.

*2.4.2. TMS Parameters*

A fixed intensity of 55% of the machine stimulator output (MSO) was set for all subjects. Stimulation was delivered in bursts of four pulses. For rhTMS, the pulse frequency corresponded to the individual theta (4-8 Hz; inter-pulse interval (IPI) = 125-250 ms) frequency extracted from the resting state EEG recording (figure 1B, top). Thus, TMS was individualized in terms of each subject’s peak theta frequency.

In the case of arrhTMS, the total duration of the burst was the same as for rhTMS, however the IPI was jittered so as to prevent the burst from having any particular frequency (figure 1B, bottom). The inter-burst delay varied from 3 to 5 seconds throughout the experiment so as to avoid carry-over effects.

***2.7. Data Analysis***

In spite of apparent usefulness of Bayesian methods, a question naturally arises: why use Bayesian statistics instead of frequentist null-hypothesis significance testing (NHST) given that the latter is still the go-to approach? In fact, Bayesian framework permits to address the very question to which every researcher seeks an answer: “what is the probability of the particular event given the data?”. Alas, NHST only permits for inverse probability which can only answer the question “what is the probability of obtaining the observed data given that the null hypothesis is true?”. The answers to the first and second questions are not equivalent (Lambert, 2018). Nevertheless, NHST has been implicitly used by the scientific community to indirectly address the first question. Bayesian inference allows to invert the probability yielded by NHST, the result being the probability of the event of interest given the observed, fixed data. Hence, we chose Bayesian hierarchical modelling as the core analysis method for this project.

This study, akin to a great number of others in psychology, used an ordinal scale to assess the construct of interest, MW. It has been pointed out, however, that the treatment of such scales as metric is still not a rare strategy among academics (Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018). Applying metric methods to ordinal variables may lead to distorted effect-size estimations, discarding of intra-individual variability and inflated Type I errors, among other issues. Also, the assumptions underlying metric methods are clearly not satisfied by ordinal variables: the scale categories are not equidistant for every subject and the resulting distribution is frequently non-normal due to responders’ unique perception of the distance between categories. Therefore, we favored the hierarchical ordered probit regression model to test the hypothesis on MW propensity. This analysis has been previously implemented in MW literature (Boayue et al., 2021; Filmer et al., 2019, 2021) and proven useful when the outcome is influenced by co-variates (e.g. total time on task). Also, this model treats the dependent variable (MW, in our case) as a rank-ordered variable and allows for the resulting hierarchically ordered data to have a non-normal distribution (Kruschke, 2014).

Similarly, BV and AE were analyzed with Baeysian hierarchical modeling: a separate model was fitted on each. In contrast to the distribution underlying the model of MW propensity, the task performance models feature Student’s-t distribution as the link function since it treats the dependent variable as a continuous, unbounded metric variable. The analysis workflow in this study consisted of three stages: model fitting, selection and non-linear hypothesis testing for the parameters of interest. The analyses were implemented with brms package in R (Bürkner, 2017) which uses Stan in the back-end.

To accommodate those readers who are more accustomed to NHST, we conducted a repeated-measures non-parametric ANOVA on the aforementioned variables which further contributed to the discussion of the results rendered by the two approaches.

*2.7.1. Model Fitting & Selection*

In our case, the modeling of MW score relies on the assumption that the variable arose from the dichotomization of a latent, unobservable variable reflecting every subject’s MW which is continuous and normally distributed (Bürkner & Vuorre, 2019). ­­­­Said distribution is divided into categories, each subject imposing unique thresholds on these categories. Hence, we assume that the response variable has a cumulative distribution, whilst the latent MW variable is normally distributed (Bürkner & Vuorre, 2019; Kruschke, 2014; Liddell & Kruschke, 2018). In brms, model fitting is done by means of Hamiltonian Monte Carlo. Given that our dataset was limited to 8 subjects, we increased the default number of iterations to 5000 per chain. For each variable of interest (MW, BV and AE) we fitted a number of models of increasing complexity.

The selection of the best model was based on the leave-one-out cross-validation (LOO-CV: Vehtari et al., 2017). For a given model and dataset of N observations, LOO-CV computes the out-of-sample prediction accuracy by fitting it on the dataset of N – 1 observations. The observation which is left out is used as an out-of-sample, new observation which is compared to the prediction of the model. This process is repeated N times. The resulting measure is the out-of-sample predictive fit (Vehtari et al., 2017):

,

where

is the leave-one-out predictive density for the data without the ith data point which was used as a new datapoint to test the prediction.

Since LOO-CV as implemented in brms relies on a LOO-approximation algorithm called Pareto-smoothed importance sampling (PSIS: Vehtari et al., 2019), the diagnostic related to this method, pareto-k, is also relevant for model selection. Pareto-k indicates how far the posterior distribution estimated without the ith datapoint is far from the distribution of the whole dataset. In cases where the disparity between the two is significantly high, the estimates of model fit are unreliable which might indicate that the model is misspecified. As a rule of thumb, if pareto-k < 0.7 for all observations, the model can be accepted. It is however desirable that pareto-k < 0.5.

Along with the selected model, we reported the following model comparison diagnostics: LOO prediction errors between models (if the selected model is not the one for which this value is zero), explained variance (Bayesian R2) and pareto-k.

*2.7.2. Non-Linear Hypothesis Testing*

Even though Bayesian statistics, *de facto*, does not involve null-hypothesis testing (Vuorre, 2020; Bürkner, 2017), we reported the results in a way that is intuitively comprehensible for anyone preferring the succinct form resembling the cut-off values of significance common for NHST. We reported the estimates returned by the method provided within the brms package: the non-linear hypothesis testing. The method allows to specify both one-way (e.g. “beta > 0”) as well as two-way (e.g. “beta ≠ 0”) contrasts.

For all hypotheses tested, the following values were reported: mean of the coefficient posterior distribution, evidence ratio in favor of a negative (ER-) or positive (ER+) direction, a credible interval (CI) including the 95% HDI as well as the posterior probability of the hypothesis being true. ER is a number which indicates the likelihood of the coefficient in question having the direction hypothesized versus the opposite direction. For instance, the hypothesis “beta > 0” could have the following estimates: mean beta = 0.77 [0.1, 0.98], ER+ = 12.13, post. prob = 0.98. The interpretation of such results could go as follows: the evidence in favor of a positive effect of the parameter in question is significant since i. all samples within the 95% CI are positive and ii. the coefficient is 12.13 times more likely to be positive than negative; based on the posterior probability, one can also assert that the probability of the true value of the coefficient being positive is 98%. Since we tested only one-way hypotheses, alpha was set to 0.025.

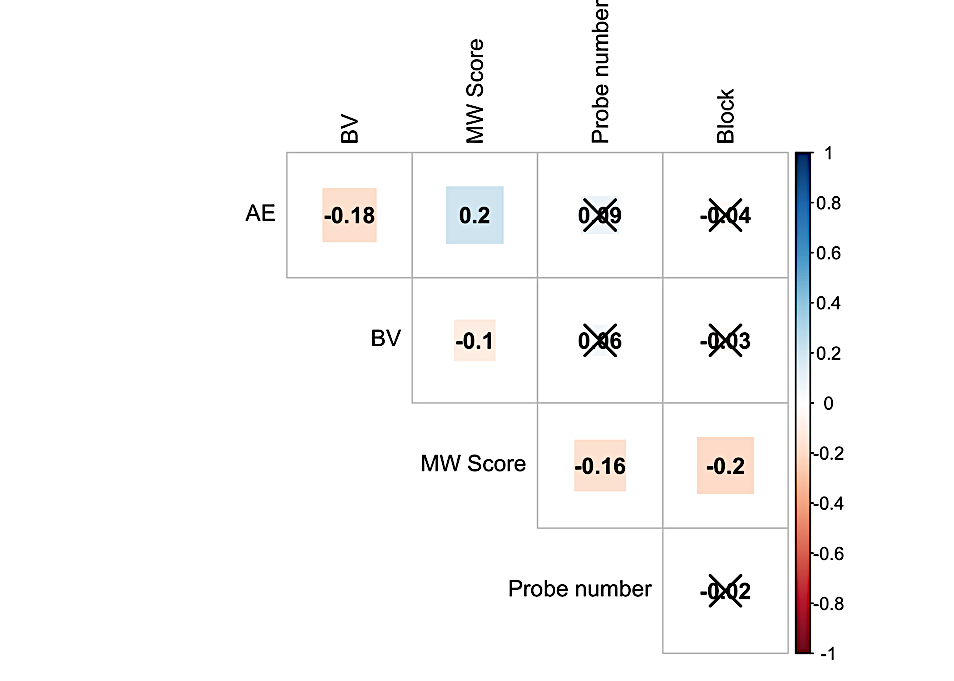
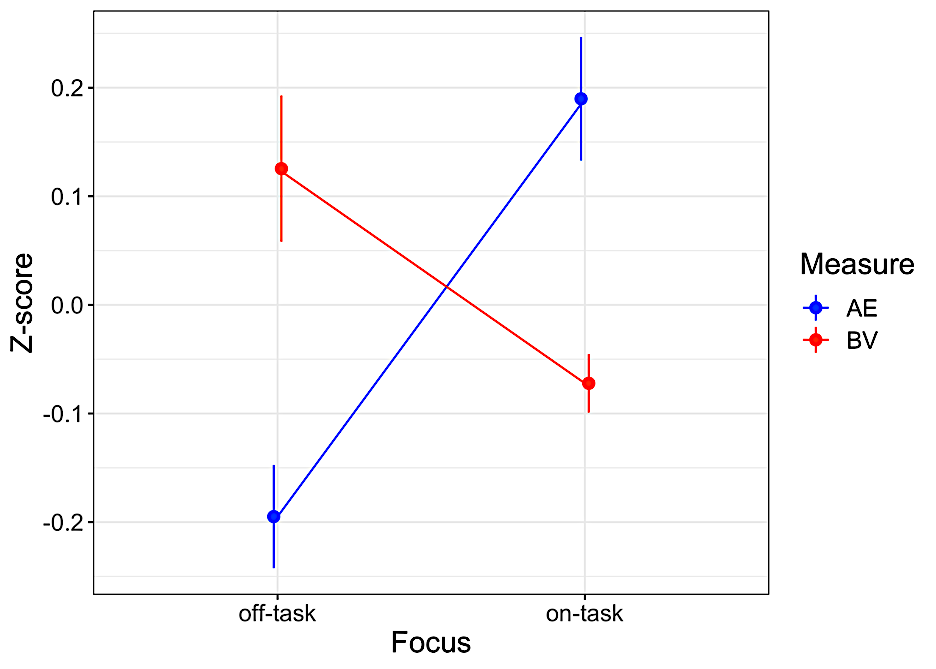
As demonstrated above, one notable advantage of non-linear hypothesis testing (as well as of Bayesian statistics more generally) is uncertainty estimates accompanying the coefficient estimates: credibility intervals indicate how many samples within the posterior distribution fall outside the 95% HDI. This straightforward measure allows for an equally straightforward interpretation of uncertainty about the true value of the coefficient.

**3. Results**

**3.1. Behavioral Pilot: Task Validation**

We conducted exploratory analyses on the pilot data whereby we replicated the interaction between AE and BV: when subjects reported being on-task (MW scores 3-4), AE was increased and BV was decreased. Conversely, when subjects were off-task, this trend was reversed: AE was decreased, and BV was increased (fig. 4).

Furthermore, we computed pairwise Spearman’s correlations for every pair of variables which were found to have significant effects on MW score in Boayue et al. (2019). In line with the interaction between AE and BV, they appeared to be negatively correlated (r = -0.18). More interestingly, MW score was also positively correlated with AE (r = 0.2) and negatively correlated with BV (r = 0.1). Time on task effects were also clearly present since block and probe number were both negatively correlated with MW score.



**Figure 6. Left:** interaction ofAE and BV depending on subject state. For on-task states: AE is increased, BV is decreased. The pattern is reversed for off-task states. This interaction is a direct replication of study 1 from Boayue et al., (2021). **Right:** Spearman’s correlation matrix for all variables potentially being linked. Probe number and block represent time on task. Non-significant (p > 0.05) coefficients are crossed out.

These findings let us further our design of the TMS-EEG experiment and allowed us to confirm that the task in question is sensitive enough to detect changes in executive control and behavioral variability. Also, the exploratory results confirmed that the link between introspective MW scores and the objective measures provided by the task is indeed present and can be further explored in the context of the preregistered TMS-EEG study.

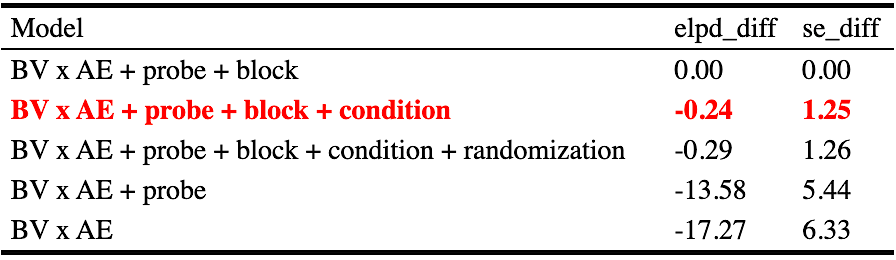
***3.2. MW Score: Hierarchical Ordered Probit Model***

*3.2.1 Model Selection*

We fitted five models of increasing complexity involving the following independent variables: BV, AE, the interaction thereof, trial and condition (baseline vs. active rhTMS vs. sham rhTMS vs. active arrhTMS vs. sham arrhTMS). Since hierarchical models allow to account for nested effects within groups or individuals, each model included intercepts nested within subjects. By introducing this effect, we explicitly acknowledge that TMS exerts an effect characterized by large intra-subject variability which cannot be ignored.

LOO diagnostics for all models fitted on the data can be viewed in table 2. The first three models had |elpd\_diff| < 4 which implies that the measure of uncertainty of model fit, *se\_diff*, is uninformative. It has been maintained that in such an instance, any model can be selected for further testing since they all have similar predictive accuracy (Sivula et al., 2020).

**Table 2.** Results of comparison of LOO-CV estimates for the fitted models. The best model is at the top.



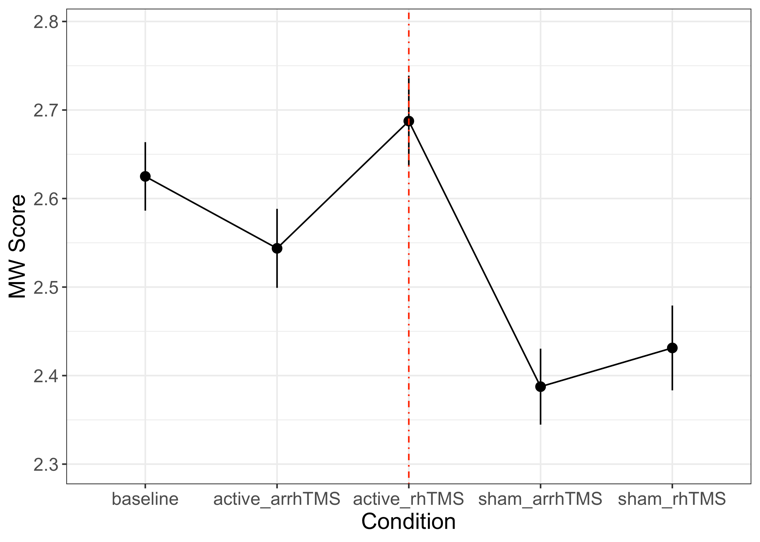
However, we favored the model which is formally considered the best as per LOO-CV:

where *condition* is a five-level variable with baseline as a reference.

While LOO-CV is a powerful statistical tool to evaluate the model’s ability to predict unobserved data, it has no way of assessing whether the model is meaningful given the parameters’ expected behavior. A series of behavioral trials of the FT-RSGT (section 3.1.) demonstrated that subjects’ task performance is characterized by an interaction of AE and BV for states off- vs. on-task. Therefore, within the model, we should expect AE and BV to have positive and negative coefficients respectively. We also anticipated the variables reflecting time on task (probe and block number) to have a negative effect on MW score. In light of these expectations, the selected model provided an adequate fit: higher BV and lower AE were indeed associated with lower MW score, whereas total time-on-task always was negatively associated with MW. Furthermore, we replicated the effect of AE x BV previously reported in Boayue et al. (2021). Coefficient estimations and corresponding CIs for each parameter can be viewed in figure 4 (left).

*3.2.2. Hypothesis Testing*

Upon visual inspection, (fig. 4, right) one can assert that rhTMS had a positive effect on MW score, i.e. during online active rhTMS subjects reported being more on-task overall. In contrast, during arrhTMS MW scores were slightly lower and during sham conditions scores were considerably lower. This variability indicated that further hypothesis testing was necessary to quantify the magnitude of TMS effects and the associated uncertainty.

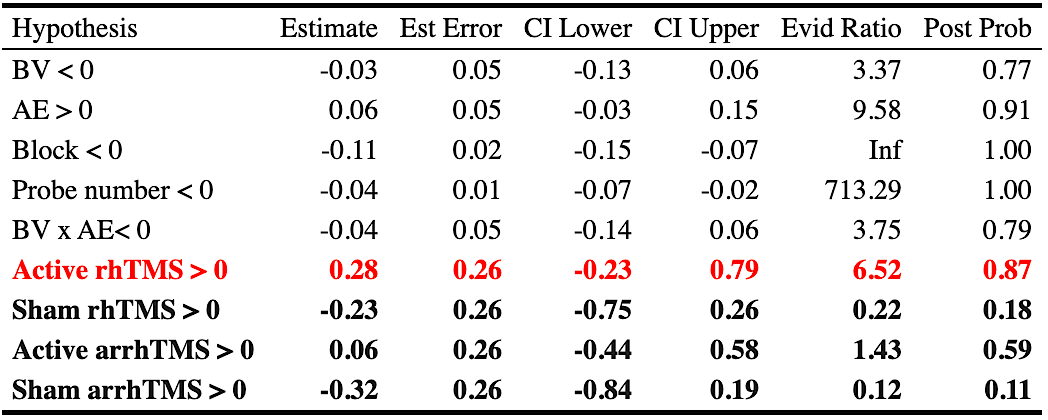
 

**Figure 7: Left:** Posterior distributions of parameter coefficients estimated for the model fitted on MW scores. Yellow circles represent posterior predictive means. Horizontal gray lines depict 95% HDIs; thick blue lines show 50% intervals. **Right:** Mean MW ± 1SE across conditions. Peak mean MW score was during active rhTMS: subjects reported being more on-task compared to other conditions. Subjects were considerably more off-task during sham conditions.

We conducted non-linear hypothesis testing (see section 2.7.2.) on all parameters of the model. As previously mentioned, probe and block number clearly have a negative effect on subjects’ self-reported concentration: zero lies outside of posterior distributions of coefficients of both parameters (fig. 4): mean β = -0.04 [-0.07, -0.02]probe, -0.11 [-0.15, -0.07]block; ER- = 713.29probe, ∞block; post. prob = 1. Posterior probability for these two parameters indicates that the effect is clearly there even though its magnitude is extremely low.

As for AE and BV, effect estimates were similarly small: β = 0.06 [-0.03, 0.15]AE, -0.03 [-0.13, 0.06]BV, -0.04 [-0.14, 0.06]AExBV. BV and BV x AE were estimated to be around 3 times more likely to be negative than positive (ER- = 3.37BV, 3.75BVxAE). Posterior probabilities also supported the negative direction even though they remained below 0.90: post. prob = 0.77BV, 0.78BVxAE. However, AE had a more definitive effect: ER+ = 9.58AE, post.prob = 0.91. Overll, the overall trend corresponded to the expected outcome, but from a purely probabilistic outlook, the link between AE, BV and MW was inconclusive and effect sizes were small.

**Table 3**. Results of non-linear hypothesis testing for each parameter. Mean coefficient estimate of active rhTMS is positive.



As for the effect of active rhTMS, we found moderate evidence in favor of its being positive: β = 0.28 [-0.23, 0.79]MW, ER+ = 6.11, post. prob. = 0.86. For other TMS conditions ER+ did not exceed 1.26 (table 3).

***3.3. Task Performance: Hierarchical Student’s-t Model***

*3.3.1 Model Selection*

Four models of increasing complexity were fitted on BV and AE metrics separately. In the case of BV, the most complex model was the best according to LOO-CV (table). However, as we pointed out above, the first three models in table 4 have similar predictive accuracy since |elpd\_diff| < 4.

**Table 4.** Models fitted on BV. The first three models have similar predictive accuracy. The model in red was selected. The models featuring variable “randomization” did not fit the data.

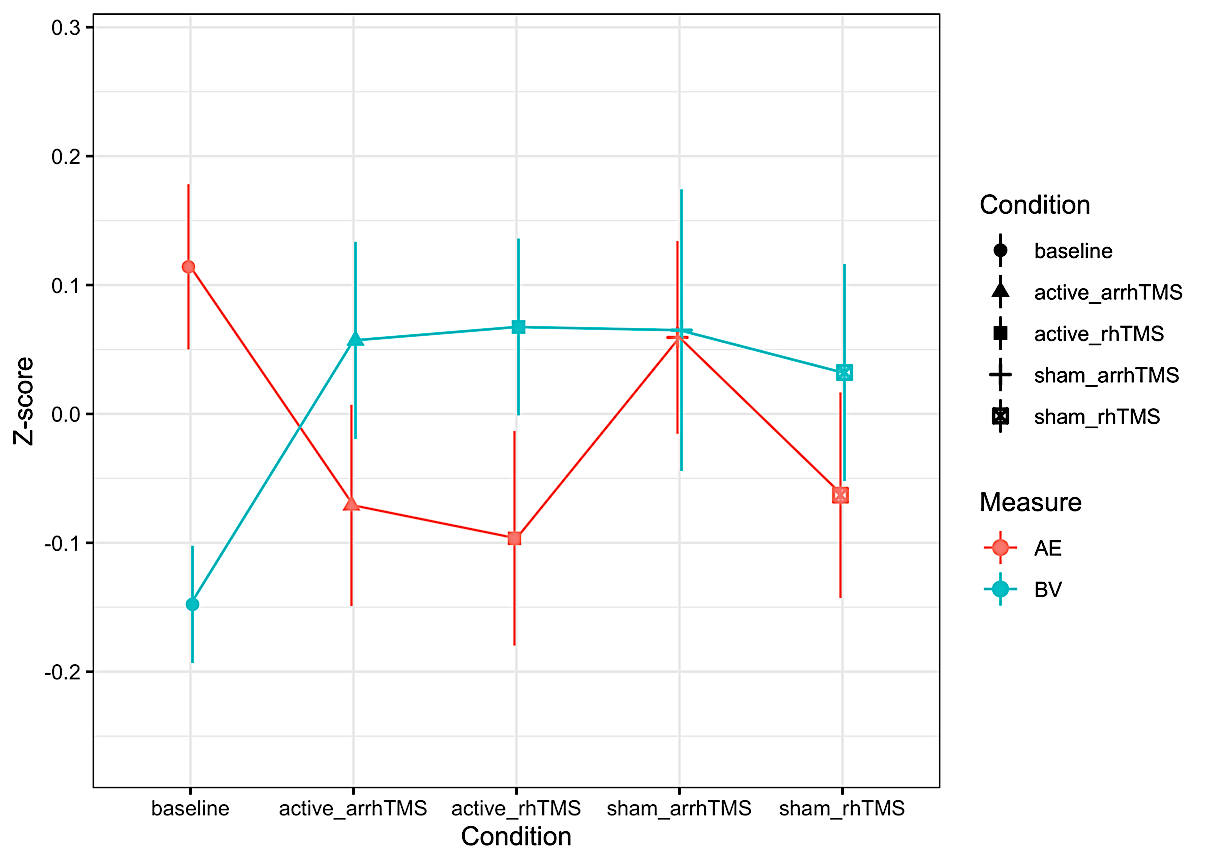


In the case of AE, for all fitted models |elpd\_diff| < 4. Therefore, we selected the same model as for BV for comparability purposes (table 5).

**Table 5.** Models fitted on AE. All |elpd\_diff| < 4. The first three models have similar predictive accuracy. The model in red was selected. The models featuring variable “randomization” did not fit the data.



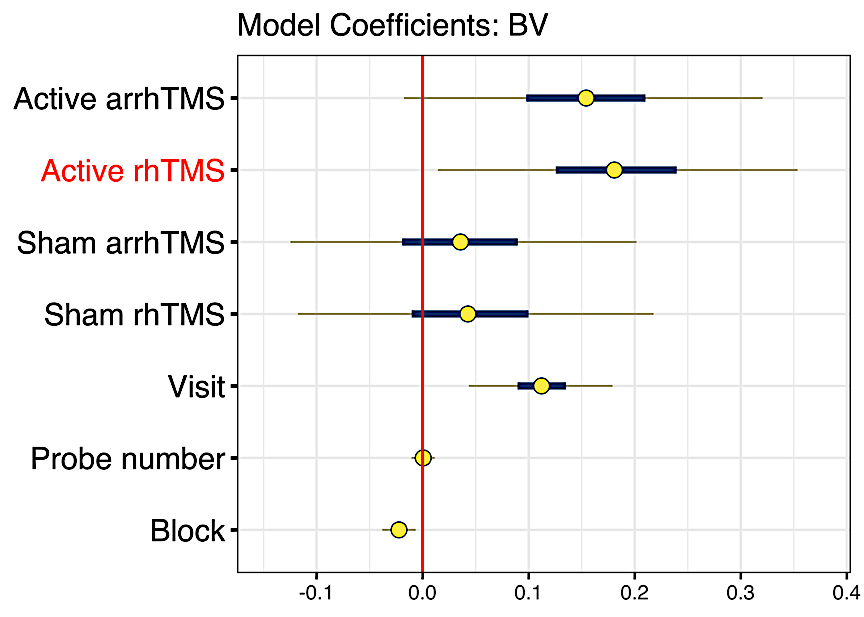
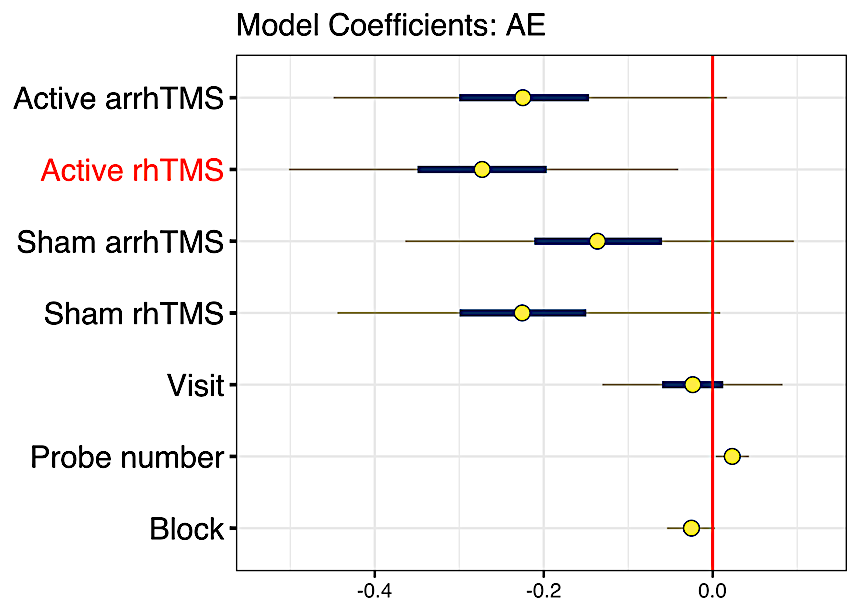
In contrast with our expectations, TMS exerted a negative effect on task performance: during all TMS conditions subjects manifested lower AE and higher BV compared to baseline.



**Figure 8.** AE and BV scores across conditions: mean ± 1SE.

In accordance with this observation, coefficients estimated for TMS effects lied on the negative side of the spectrum for AE (fig. 5, left) and the trend was reversed for BV (fig. 5, right). Moreover, active rhTMS undoubtedly rendered task performance worse since for both metrics: zero was outside of the corresponding posterior distributions. The distributions of other TMS conditions had exerted a similar effect on task metrics even though for some samples the coefficient was zero.

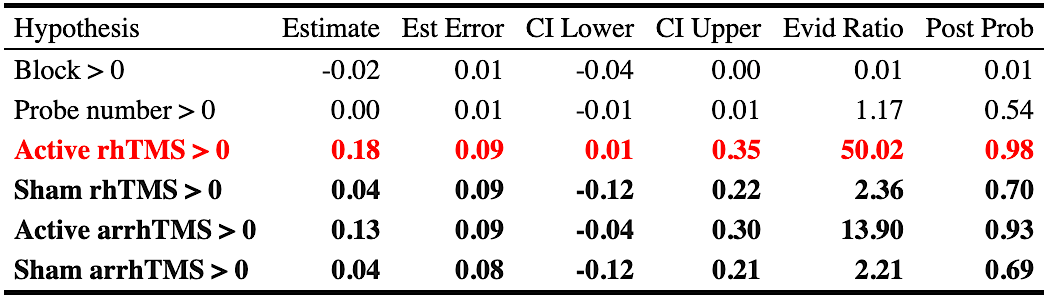
Figures 8 and 9 alone were enough to convince us that the direction of the effect of the stimulation was the opposite of what we hypothesized for AE and BV. Hence, since Bayesian statistics allows to test “null”[[1]](#footnote-1) hypotheses, we did hypothesis testing in this reversed direction, i.e. positive for BV and negative for AE. The exact posterior estimates can be viewed in tables 6 and 7.



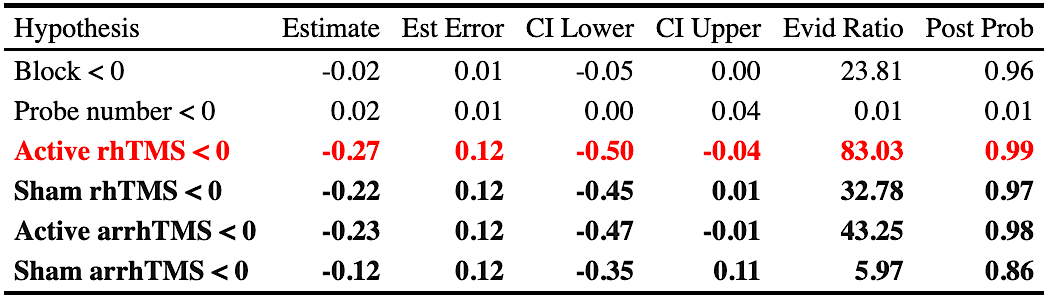
**Figure 9.** Posterior distributions of parameter coefficients estimated for the models fitted on AE and BV. Yellow circles represent posterior means. Horizontal gray lines depict 95% HDIs; thick blue lines show 50% intervals.

We found strong evidence in favor of the negative effect of active TMS regardless on task metrics regardless of rhythmicity: AE was lower and BV was higher during TMS sessions compared to baseline. The strongest evidence of that effect pertains to active rhTMS: β = 0.18 [-0.01, 0.35]BV, -0.027 [-0.50, -0.04]AE;ER-AE = 84.47, ER+BV = 56.47; post.prob = 0.99AE, 0.98BV. Interestingly, time on task effects were insignificant for both measures.

**Table 6.** Results of non-linear hypothesis testing for the model of BV. All TMS conditions increased BV: subjects’ finger tapping was less in sync with the metronome during TMS sessions compared to baseline. Time-on-task effects were negligible.



**Table 6.** Results of non-linear hypothesis testing for the parameters of the model fitted on AE. All TMS conditions decreased AE: subjects were more predictable when alternating between left and right taps during TMS sessions compared to baseline. The strongest effect was exerted by active rhTMS. Time-on-task effects were negligible.



***3.3. ANOVA: Kruskal-Wallis Test***

Since the assumption of normality was violated for all dependent variables (Shapiro-Wilk test: W = 0.57-0.94, p < 0.001, fig. SX), we conducted a non-parametric version of ANOVA, the Kruskal-Wallis test. When performed on MW scores, the test indicated that MW propensity differed significantly across TMS conditions: H (4) = 15.76, p = 0.003. A *post-hoc* one-way multiple comparisons test (Siegel & Castellan, 1981) revealed that the difference driven by sham vs. active rhTMS contrasts, with only active rhTMS vs. sham arrhTMS contrast surviving the correction for multiple comparisons (Wilcoxon test: W = 41168, r = 0.18, p < 0.0001, Ncomparisons  = 10).

In the case of AE, the difference between conditions was also significant: H (4) = 9.63, p = 0.047. The result was mainly driven by the baseline vs. active rhTMS contrast: W = 22138, r = 0.13, p = 0.009, Ncomparisons = 4 (two-way). However, this contrast did not reach the corrected level of significance when the number of comparisons increased to 10. The same testing procedure did not yield significant results for BV.

**4. Discussion**

In the present study we combined the finger-tapping random sequence generation task (FT-RSGT) and a repeated-measures online TMS protocol. The aim of the project was to probe the resulting data for a causal link between the propensity to mind-wander and theta-band entrainment. The latter was implemented by means of rhythmic TMS: every 3 to 5 seconds during the task, subjects were administered a burst of four pulses at individual theta frequency. Each subject underwent 110 minutes of the finger-tapping random sequence generation task (FT-RSGT) in two visits. To control for the rhythmicity of the stimulation and placebo effects separately, we designed four control conditions: baseline (no TMS), active arrhythmic, sham arrhythmic and sham rhythmic TMS, active rhythmic TMS being the condition of interest. Thus, at each visit, subjects were exposed to all five conditions.

In the context of the TMS-EEG study, we hypothesized a positive effect of active rhTMS on MW scores, i.e. we expected the subjects to report being more on-task during the active rhythmic condition compared to baseline. Furthermore, we hypothesized a similar positive effect on task performance: we anticipated lower BV and higher AE during active thTMS compared to baseline. To control for the rhythmicity of the stimulation and placebo effects separately, we designed three TMS control conditions: active arrhythmic, sham arrhythmic and sham rhythmic TMS. Thus, in total, the subjects were exposed to five conditions.

Given the relative novelty of the FT-RSGT, we preliminarily verified its content validity in the context of a pilot study. Subjects were instructed to sequentially press the left or right key in a set rhythm in such a manner that i. the rhythm of the tapping is perfectly synchronized with the metronome and ii. each key press, left or right, is maximally unpredictable for the external observer. FT-RSGT yields two outcome measures: approximate entropy (AE) and behavioral variability (BV), the former reflecting executive control. Whist AE measures the irregularity of the left-right press sequency, BV quantifies the temporal deviation of the tapping rhythm from the metronome. The results of exploratory analyses on the pilot data demonstrated that the outcome measures were correlated with MW score as expected: subjects produced more regular sequences of left-right presses and were more out of sync with the metronome when MW score was lower. A direct replication of the trend uncovered by Boayue et al. (2021), this result allowed us to proceed with further experimentation involving TMS.

We used Bayesian hierarchical models as the primary analysis method: from a number of potential models fitted on MW, AE and BV scores we chose those with the best LOO diagnostics provided they featured the condition variable. As for MW, analyses provided weak evidence in favor of the first hypothesis: whilst the coefficient estimate associated to rhythmic TMS did indicate that subjects reported being more on-task during the said block, the posterior probability of a positive effect did not exceed 0.77. From these estimates we can conclude a number of things. Firstly, the hypothesized effect is present, although the uncertainty surrounding its true value is high. That is hardly surprising given the limited sample size of only eight subjects.

Contrary to the hypotheses of lower behavioral variability and higher executive control during active rhTMS block, we found a completely reversed pattern: task performance was worse for all TMS blocks compared to baseline. This led us to change the direction of the hypothesized effect so as to make the estimates more intuitively interpretable. The resulting estimates strongly supported the deleterious effect of active rhTMS on both task metrics: subjects were less in sync with the metronome and more predictable in their left-right tapping patterns. Notably, the posterior probability of these effects was close to 1 for both metrics which implies minimal uncertainty about the true direction of the effect of TMS. Sham and active arrhythmic conditions also appeared to be associated with worse task performance, albeit with greater uncertainty estimates.

In addition to Bayesian analyses, we conducted a non-parametric ANOVA, Kruskal-Wallis test, to compare the insight provided by both Bayesian and NHST approaches. Unsurprisingly, compared to hierarchical models, the results of the non-parametric test and post-hoc tests were not perfectly aligned, albeit comparable. Thus, MW scores indeed appeared to differ across conditions, although only the comparison between sham arrhTMS and active rhTMS survived the correction for multiple comparisons. From these results indicate that MW score was indeed the highest during active rhTMS among all conditions which is in line with the results produced by Bayesian methods.

The same testing procedure applied to AE similarly revealed a pattern which suggested the strongest effect of active rhTMS on task performance: active rhTMS compared to baseline significantly decreased AE. However, significant difference was detected for BV scores.

These findings taken together paint a complex picture of effects of TMS on MW propensity. We have seen that the interpretation of results is heavily dependent on the statistical framework: whilst Bayesian analyses provide posterior probability estimates of effects, NHST mainly leads us to accept or reject the null hypothesis, giving us little information about the alternative hypothesis given the data. This is the main reason for our usage of Bayesian hierarchical models: with only 8 subjects we managed to assess the magnitude and the direction of the hypothesized effects, evaluate the uncertainty intrinsic to these values and to do reverse hypothesis testing when faced with the necessity to reject the hypothesis. More concretely, while values returned by the non-parametric procedure do not give rise to a drastically different interpretation of the effects, it does not provide the same amount of information. In light of the richness of insight offered by the Bayesian framework and its flexibility in the context of limited samples, we argue in favor of its advent in psychological sciences.

**REFERENCES**

Axelrod, V., Rees, G., Lavidor, M., & Bar, M. (2015). Increasing propensity to mind-wander with transcranial direct current stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3314–3319.

Axelrod, V., Zhu, X., & Qiu, J. (2018). Transcranial stimulation of the frontal lobes increases propensity of mind-wandering without changing meta-awareness. *Scientific Reports*, *8*(1), 15975.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *51*(4), 819–852.

Boayue, Nya M., Csifcsák, G., Kreis, I. V., Schmidt, C., Finn, I., Hovde Vollsund, A. E., & Mittner, M. (2021). The interplay between executive control, behavioural variability and mind wandering: Insights from a high-definition transcranial direct-current stimulation study. *The European Journal of Neuroscience*, *53*(5), 1498–1516.

Boayue, Nya Mehnwolo, Csifcsák, G., Aslaksen, P., Turi, Z., Antal, A., Groot, J., Hawkins, G. E., Forstmann, B., Opitz, A., Thielscher, A., & Mittner, M. (2019). Increasing propensity to mind-wander by transcranial direct current stimulation? A registered report. *The European Journal of Neuroscience*, *51*(3), 755–780.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101.

Callard, F., Smallwood, J., Golchert, J., & Margulies, D. S. (2013). The era of the wandering mind? Twenty-first century research on self-generated mental activity. *Frontiers in Psychology*, *4*, 891.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421.

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.

Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*, *19*(4), 188–195.

Duecker, F., & Sack, A. T. (2015). Rethinking the role of sham TMS. *Frontiers in Psychology*, *6*, 210.

Filmer, H. L., Griffin, A., & Dux, P. E. (2019). For a minute there, I lost myself … dosage dependent increases in mind wandering via prefrontal tDCS. *Neuropsychologia*, *129*, 379–384.

Filmer, H. L., Marcus, L. H., & Dux, P. E. (2021). Stimulating task unrelated thoughts: tDCS of prefrontal and parietal cortices leads to polarity specific increases in mind wandering. *Neuropsychologia*, *151*, 107723.

Gouraud, J., Delorme, A., & Berberian, B. (2018). Out of the Loop, in Your Bubble: Mind Wandering Is Independent From Automation Reliability, but Influences Task Engagement. *Frontiers in Human Neuroscience*, *12*, 383.

Groot, J. M., Boayue, N. M., Csifcsák, G., Boekel, W., Huster, R., Forstmann, B. U., & Mittner, M. (2020). Probing the neural signature of mind wandering with simultaneous fMRI-EEG and pupillometry. *NeuroImage*, *224*, 117412.

Hoffmann, F., Banzhaf, C., Kanske, P., Bermpohl, F., & Singer, T. (2016). Where the depressed mind wanders: Self-generated thought patterns as assessed through experience sampling as a state marker of depression. In *Journal of Affective Disorders* (Vol. 198, pp. 127–134). https://doi.org/10.1016/j.jad.2016.03.005

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*(7), 614–621.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932.

Klinger, E., & Cox, W. M. (1987). Dimensions of Thought Flow in Everyday Life. *Imagination, Cognition and Personality*, *7*(2), 105–128.

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2016). Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral Cortex* . https://doi.org/10.1093/cercor/bhw029

Lambert, B. (2018). *A Student’s Guide to Bayesian Statistics*. SAGE.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008) [Review of *Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008)*]. *Psychological Bulletin*, *136*(2), 188–197; discussion 198-207.

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: a review. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *67*(1), 11–18.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(6), 2297–2301.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!”: performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmöller, J., Carpenter, L. L., Cincotta, M., Chen, R., Daskalakis, J. D., Di Lazzaro, V., Fox, M. D., George, M. S., Gilbert, D., Kimiskidis, V. K., Koch, G., Ilmoniemi, R. J., Lefaucheur, J. P., Leocani, L., … Hallett, M. (2021). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert Guidelines. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *132*(1), 269–306.

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really?,. *Consciousness and Cognition*, *66*, 74–78.

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: linking mind wandering and behavioral variability. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(1), 1–5.

Siegel, S., & Castellan, N. J. (1981). Nonparametric Statistics for the Behavioral Sciences. *McGraw-HiU Book Company, New York*.

Sivula, T., Magnusson, M., Matamoros, A. A., & Vehtari, A. (2020). Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/2008.10296

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487–518.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology* , *89 ( Pt 1)*, 77–101.

Turnbull, A., Wang, H. T., Murphy, C., Ho, N. S. P., Wang, X., Sormaz, M., Karapanagiotidis, T., Leech, R. M., Bernhardt, B., Margulies, D. S., Vatansever, D., Jefferies, E., & Smallwood, J. (2019). Left dorsolateral prefrontal cortex supports context-dependent prioritisation of off-task thought. *Nature Communications*, *10*(1), 3816.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2019). Pareto Smoothed Importance Sampling. In *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1507.02646

Vuorre, M. (2020). *Sometimes I R: How to calculate contrasts from a fitted brms model.* Retrieved from https://mvuorre.github.io/posts/2020-02-06-how-to-calculate-contrasts-from-a-fitted-brms-model/

1. There is no such thing as null hypothesis in Bayesian statistics per se. It is more practical to think of Bayesian methods as tools to test any hypothesis, even if in NHST it is considered as null. [↑](#footnote-ref-1)