UiT The Arctic University of Norway

Faculty of health sciences, department of psychology

Replicating the effect of brain stimulation on mind wandering: A pre-registered study

Andreas Alexandersen

Master's thesis in Psychology, PSY-3900 - Tromsø, May 2022

Preface

This project is part of a larger initiative to investigate mind wandering, started by Professor Matthias Mittner and the cognitive neuroscience research group at the University of Tromsø (UiT). Consequently, the idea behind this master project and its extension the Forskerlinje project: "The effect of transcranial direct current stimulation on the interplay between executive control, behavioural variability and mind wandering: a registered report" is a collaboration with Prof. Mittner, and associate professor Gábor Csifcsák. For the past three years I have worked closely together with my supervisors Matthias Mittner, Gabor Csifcsak and Ph.D. student Josephine Groot, to complete these two projects which ended up in one accepted manuscript at NeuroImage: Reports with the focus on the neural mechanisms behind mind wandering, and one master thesis focusing on the replication aspect of the project.

It started with me knocking on Matthias' door and telling him about my interest in cognitive neuroscience, asking if he had any projects going on and wanted to supervise me as a master student. About a year later I also got accepted into the Forskerlinje program. Since then, I have been working together with the cognitive neuroscience group and my supervisors to complete this project. Together with my supervisors we agreed on study design, while I was in charge of setting up the lab equipment and data collection. I assisted Matthias Mittner in coding of the task script as well as the data pre-processing and statistical analysis.

First and foremost, I want to thank my supervisors for the great work environment they provided during my time in the cognitive neuroscience group. I also want to thank my fellow master and Forskerlinje students, the Ph.D. candidates at IPS for their valuable input and all my friends and family who supported me throughout my studies.

Andreas Alexandersen
(Student)

Matthias Mittner
(Supervisor)

Replicating the effect of brain stimulation on mind wandering: A pre-registered study

Andreas Alexandersen

Supervisor: Matthias Mittner

PSY-3900

Institute for Psychology, UiT The Arctic University of Norway

Abstract

Mind wandering (MW) is a common mental phenomenon. Despite this, there is still much we don't know about this pervasive mental state. Transcranial direct current stimulation (tDCS) has been proposed to be able to modulate mind wandering propensity, but a large variability in results paints an inconclusive picture in the current brain stimulation literature, and a satisfactory conclusion is still lacking. Recently, a study by Boayue et al. (2020) reported to successfully reduce mind wandering using high-definition transcranial direct current stimulation (HD-tDCS) over the dorsolateral prefrontal cortex, providing preliminary evidence of the efficacy of HD-tDCS in modulating MW. The current thesis introduces the topic of replicating this elusive effect of non-invasive brain stimulation in depth, as well as reporting a high-powered, pre-registered direct replication attempt of the effect found by Boayue et al. (2020). Additionally, the results of investigating MW with a finger-tapping random sequence generation task that draws heavily on executive resources are reported. We failed to replicate the original effect of reducing MW during HD-tDCS, and in a meta-analytic approach, when the data was combined with Boayue et al. (2020) the original effect of HD-tDCS reducing MW disappeared. These findings and potential problems of brain stimulation studies (in particular their low replicability) and their implications are reviewed and discussed.

Abstrakt

Tankevandring er et vanlig mentalt fenomen. Til tross for at tankevandring er vanlig i hverdagen, er det fortsatt mye vi ikke vet om denne mentale tilstanden. Transkraniell likestrøms stimulering (tDCS) har blitt foreslått å kunne modulere mengden individer tankevandrer i en laboratoriesetting, men et stort avvik i resultatene maler et usikkert bilde i dagens hjernestimuleringslitteratur, og en tilfredsstillende konklusjon mangler fortsatt. Nylig rapporterte en studie av Boayue et al. (2020) at de klarte å redusere tankevandring ved hjelp av høy-definisjon transkraniell likestrømstimulering (HD-tDCS) over den dorsolaterale prefrontale kortex, noe som gir foreløpige bevis på effektiviteten til HD-tDCS i modulering av tankevandring. Denne oppgaven introduserer temaet om å reprodusere effekten av hjernestimuleringsstudier, i tillegg til å rapportere ett forhåndsregistrert direkte replikasjonsforsøk av effekten funnet av Boayue et al. (2020). I tillegg rapporteres resultatene av å undersøke tankevandring med en tilfeldig sekvensgenereringsoppgave basert på fingertapping som krever tung bruk av eksekutive ressurser. Vi klarte ikke å gjenskape den opprinnelige effekten av å redusere tankevandring ved bruk av HD-tDCS. I tillegg ble en meta-analytisk tilnærming gjennomført, og utvalget ble kombinert med utvalget fra Boayue et al. (2020), noe som førte til at den opprinnelige effekten at HD-tDCS reduserte tankevandring, forsvant. Disse funnene og potensielle problemer med hjernestimulering (og den lave reproduserbarheten av slike studier) og deres implikasjoner blir gjennomgått og diskutert.

Replicating the effect of brain stimulation on mind wandering: A pre-registered study

In the past decade, altering human cognition has become an increasingly popular topic. In everyday life, doctors, psychologists and private individuals are using a plethora of methods to alter human brain activity (e.g., deCharms, 2007; Han et al., 2011; Lefaucheur et al., 2014). For example, humans have attempted to modulate brain activity in multiple forms and for multiple purposes, ranging from medicinal or recreational drug use (Brick & Erickson, 2012) and cognitive therapy (Cuijpers et al., 2013), to mindfulness sessions and meditation (Chiesa & Serretti, 2009). Recently, a new research method has seen an increase in popularity, non-invasive brain stimulation (NIBS), where the goal is to apply an electrical or magnetic field to the brain non-invasively, to change brain function from the outside (Kuo & Nitsche, 2012). More specifically, NIBS aims to achieve the possibility of altering certain cognitive functions without the use of drugs, as an alternative to the current therapeutic options (Brunoni & Fregni, 2011; though see; Tortella et al., 2014). The idea of being able to substitute drugs to circumvent common problems with pharmaceutical interventions such as absorption and side effects or speed up the process of therapy by using NIBS is a promising development in the field of neuromodulation.

One of the cognitive functions that has been a popular research topic in psychology is attention (Driver, 2001; Ocasio, 2011), recently attention has also become popular topic to research with NIBS methods (Rubio et al., 2016). While attention is a naturally interesting topic to study, attention is also being researched due to its implication in psychological disorders such as ADHD (Luo et al., 2019; Tarver et al., 2014; Westwood et al., 2021), and mood disorders such as anxiety and depression (Ottaviani & Couyoumdjian, 2013). One of the key contributors to mood disorders such as anxiety and depression is shifting attention away from external stimuli to internally generated (negative) thoughts (Ottaviani & Couyoumdjian, 2013). Unintentional shifts in attention are not only associated with

pathology, the feeling of realizing your thoughts no longer are focused on the task at hand is very common (Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2006). It is estimated that we spend as much as half of the time we are awake occupied with thoughts unrelated to the current task at hand, a condition that is often called mind wandering (MW; Antrobus et al., 1970; Kane et al., 2007; McVay et al., 2009). Additionally, researchers have confirmed that shifting the attentional focus away from the task at hand reduces task performance (Yanko & Spalek, 2013). Although these types of attentional shifts or MW are often inconsequential in everyday tasks such as reading, it can be more detrimental during tasks such as driving or in aviation where attentional lapses can have severe consequences (Yanko & Spalek, 2013).

Reducing MW to increase task focus or performance by electrically stimulating the brain could be desirable for practicians to supplement current treatment of mood disorders, as well as potentially increasing efficiency in task that requires focus (studying, reading). Unfortunately, results from research on whether it is possible to modulate attention with NIBS are still unclear, though not for the lack of trying (Axelrod et al., 2018; Boayue et al., 2019; Coulborn et al., 2020). The current literature paints a picture of conflicting results, with some studies finding strong effects of NIBS on MW and others finding evidence against that. This situation urgently calls for systematic and thorough investigations with more sophisticated methods and rigorous study designs.

Transcranial direct current stimulation

One of the most popular NIBS techniques used to investigate attention and MW is transcranial direct current stimulation (tDCS; Nitsche et al., 2008). TDCS operates by applying a low-intensity current to the brain that induces an electric field in the cortex. This electric field can either depolarize or hyperpolarize the neuronal membrane potential of cortical pyramidal neurons (Huang et al., 2017; Nitsche & Paulus, 2001). This (de-)

polarization can alter cortical excitability, thereby either increasing or decreasing activity depending on the polarity of the flow of the current through the targeted neurons (Purpura & McMurtry, 1965). Research has even found evidence of neuroplastic effects when tDCS is used over longer stimulation periods (Lefaucheur et al., 2014; Nitsche & Paulus, 2001). The possibility of creating lasting effects to either depolarize or hyperpolarize neuronal resting potential makes tDCS a versatile tool for modulating brain activity and in turn interfering with mechanisms which is responsible for the phenomenon of MW.

Mind wandering

While research on MW has increased in the past decade (Christoff et al., 2018; Seli et al., 2018), we still do not know exactly how MW occurs, making it hard to target this phenomenon with tDCS. However, there literature suggests that executive functions (EF) and executive resources (ER) are involved. The term executive functions have had many definitions over the past decades (Karr et al., 2018). Early definitions of executive functions were described as frontal lobe functioning (Pribram, 1973), control over lower-level cognition (Baddeley & Hitch, 1974) and later as an umbrella term for cognitive control (Friedman & Robbins, 2022).

Different distinctions have been made to attempt to pick apart the concept of EF. As an example, Miyake et al. (2000) proposed three distinct categories of EF; inhibition of prepotent responses, updated of working memory representation, and mental set-shifting. In a recent study on MW and EF, researchers investigated how these three distinctive types of EF was related to MW. They found that MW disrupted performance in tasks requiring inhibition of prepotent responses and updating of working memory representation, but not mental set-shifting (Kam & Handy, 2014). This suggest that switching between task engagement and MW might not always recruit all types of EF, depending on task demands. Indeed, not

everyone agrees on the exact definition of EF, and there may be some discrepancy depending on context and field of interest (e.g., Karr et al., 2018). Nevertheless, there seems to be a consensus that EF is a placeholder term for a variety of complex processes that are necessary for problem-solving and higher cognition, which is operated by the frontal lobes (Alvarez & Emory, 2006), and that ER are often referred to as the availability of these functions (Alvarez & Emory, 2006; Gross & Grossman, 2010).

While there still is disagreement amongst researchers on how EF are related to MW, there are currently two competing hypotheses: The executive function failure (EFF) hypothesis, and the executive function use (EFU) hypothesis (McVay & Kane, 2010; Smallwood & Schooler, 2015; Watkins, 2008). The EFF hypothesis postulates that MW occurs as a failure of executive control. This view does not assume that MW use the same ER as executive control, but instead that MW can be prevented by using the executive control system. The EFF postulates that MW occurs when the executive control system fails (which in turn can be caused by lack of available ER; McVay & Kane, 2010). The EFU hypothesis, on the other hand, postulates that the task and MW share the same resources and therefore the executive resources can be allocated and shared between the task and MW (Smallwood & Schooler, 2015; Watkins, 2008).

While both hypotheses explain how task performance decreases when MW occurs, they predict opposite outcomes depending on whether the available ER are increased or decreased. As an example, the EFF views MW as a result of failing to have the required resources for executive control demands, and therefore increasing the available ER would help to protect against intrusive thoughts and result in reduced MW. On the other hand, the EFU hypothesis would predict the opposite: If MW and task demands share the same ER, increasing the amount of available ER would enable the participant to satisfy task demands as well as engaging in MW, consequently resulting in increased MW.

Investigating executive functions and MW in the lab

To investigate executive function and MW in a lab setting, many different cognitive tasks have been invented and adapted over the years. The most prominent and widely used task to investigate MW is the sustained attention to response task (SART; Smallwood & Schooler, 2006). The SART is a Go/NoGo task which requires the participant to respond to certain target stimuli, creating a need to keep sustained attention to the task to perform well. The task works by presenting stimuli (numbers 0-9) in succession, and the participant is required to respond to all stimuli except a fixed target stimulus (usually the number 3). The SART produces several behavioral indices of MW. Firstly, incorrect responses to the NOGO stimulus (commission errors) is the primary outcome of the SART, and is used as an indication of lapses in attention (Seli et al., 2013). Secondly, failure to respond to the GO stimuli (often called omissions), is a common human error which can be interpreted as not engaging in the task and therefore reflects lapses in attention or engagement in MW (Johnson et al., 2007). The SART also captures other indices of MW, such as variance of response timers (RTs), where it is common to observe very short or very long RTs during episodes of MW (Cheyne et al., 2009), as well as speeding of RTs on GO trials directly after NOGO errors (Cheyne et al., 2009). Additionally, responses to GO trials which are too fast to be responses to the GO stimuli, are indices of anticipation and can be interpreted as result of autopilot behavior due to the set interval time of the SART (Cheyne et al., 2009). Altogether, behavioral indices captured by the SART can therefore be interpreted as a reflection of EF use (Cheyne et al., 2009; Seli et al., 2013).

However, the SART has received some criticism recently, namely that due to the low temporal resolution of the task (target stimuli occur only rarely), the task is very monotonous and might not require as much executive function as previously suggested (Boayue et al., 2020). While the SART is robust and versatile, new cognitive tasks have been developed to

specifically investigate the intricate relationship between EF and MW. A recent study published evidence of a finger-tapping random sequence generation task (FT-RSGT), which combined the classical finger tapping task (Kucyi et al., 2017; Seli et al., 2013) and a random number generation task (Baddeley et al., 1998; Towse, 1998), showing promising results in being able to capture behavioral manifestation of MW episodes with a high temporal resolution (Boayue et al., 2020). The idea behind the FT-RSGT is that creating random sequences draws heavily on executive resources, and therefore the onset of MW should be reflected in decreased task performance when the executive resources are being used up by MW.

When it comes to the distinction of which types of EF the FT-RSGT requires, in accordance with the distinctions proposed by Miyake et al. (2000), the FT-RSGT requires inhibition of prepotent responses as well as updating of working memory. Inhibition of prepotent responses is required to suppress an inherent tendency for stereotypical patterns and updating of working memory is required to keep track of the previous sub-sequences generated. More uncertain is the requirement for mental-set shifting, and it is also suggested that the difficulty of the task also plays a role in to which extend EF is recruited (Kam & Handy, 2014). Nevertheless, the FT-RSGT can be a powerful tool for researchers to investigate the interplay between EF and MW. This was later supported by evidence of neural signatures captured with functional magnetic resonance imaging (fMRI), and pupillometric measurements (pupillometry) during the FT-RSGT (Groot et al., 2022).

When it comes to the neural basis of MW, evidence suggests that the posterior parietal cingulate, the medial prefrontal cortex and the medial temporal lobes, also known as the default mode network (DMN) is involved (Christoff et al., 2016). The DMN is most active after external stimuli are gone, and it is proposed that the DMN is a supplier of content for MW episodes (Andrews-Hanna et al., 2010; Dixon et al., 2018; Fox et al., 2013, 2015; Kam et

al., 2022). Together with the DMN, the frontoparietal control network (FPCN) has also been implicated to partake in MW (Christoff et al., 2016; Fox et al., 2015; Kam et al., 2022). The FPCN is a key agent in executive control processes (Christoff et al., 2016; Kam et al., 2022), and it is therefore suggested that the FPCN and DMN work together to actively select and provide content which gives rise to the experience of MW (Smallwood et al., 2012; Spreng et al., 2010).

Using tDCS to modulate MW

To investigate whether tDCS can modulate MW, multiple brain regions have been highlighted as potential targets. Since MW is related to EF, the dorsolateral prefrontal cortex (DLPFC) as a part of the FPCN is a popular target due to the FPCN's involvement in control and coordination (Christoff et al., 2016). The DLPFC is also a popular target due to its superficial location and accessibility (Seibt et al., 2015). Multiple studies have attempted to modulate MW by applying tDCS over the left DLPFC (Axelrod et al., 2015, 2018; Boayue et al., 2019, 2020). Unfortunately, a satisfactory conclusion is lacking as to whether non-invasive brain stimulation can modulate MW. While early studies found very large effects of tDCS increasing MW when stimulating the DLPFC (Axelrod et al., 2015), as well as reducing MW by stimulating the right inferior parietal lobule (Kajimura & Nomura, 2015), the effects found in these initial studies were based on small sample sizes, and the final analysis pipeline was not registered beforehand. Later these effects have failed to replicate in a larger sample (Boayue et al., 2019; Coulborn et al., 2020; though see Axelrod et al., 2018; Csifcsak et al., 2019).

Since then, multiple different stimulation montages, stimulation intensities and target regions have been used to replicate the preliminary findings of tDCS modulating MW with varying results (e.g., Boayue et al., 2020; Chou et al., 2020; Coulborn et al., 2020; Filmer et

al., 2019, 2021). While there are some studies who claim to successfully modulate MW with tDCS (Axelrod et al., 2018; Boayue et al., 2020; Filmer et al., 2019, 2021), some of these reports find opposite effects when stimulating the same brain region with the same polarity (anodal stimulation of the DLPFC; increased MW; Axelrod et al., 2018; decreased MW; Boayue et al., 2020). This calls into question the robustness of the original results.

While some researchers have showed their enthusiasm for the promise of modulating MW with tDCS, going as far as calling it a new era in MW research (Broadway et al., 2015), other researchers are more skeptical of drawing any definite conclusions about the efficacy of tDCS on cognitive functions altogether (Horvath et al., 2015; Tremblay et al., 2014). This disagreement, and uncertainty in the current literature calls for more direct replications before any effect of tDCS on MW should be accepted.

Replication and researchers' degrees of freedom

While NIBS studies, like all other experiments, suffer from the variance from small differences across labs and tasks, there has also been reported problems with brain-stimulation specific sources of variance such as differences in individual anatomy and stimulation montages (Horvath et al., 2015). Data collection in NIBS studies are often costly and time consuming, resulting in smaller sample sizes. In addition to the problems with small sample sizes and high between subject variance, another problem for replication in brain stimulation studies, is the use of flexible analysis designs, which often is referred to as "researcher degrees of freedom" (Simmons et al., 2011). This term refers to the flexibility researchers have when designing the analysis plan post data collection, as well as when reporting the results. Extensive use of analytical flexibility has been suggested to be the biggest reasons why a lot of psychological research fails to replicate (Asendorpf et al., 2013, 2016). When designing a study, decisions are made about data collection, exclusion, and inclusion of

variables and/or data, sample size, hypothesis formulation, which results to report and many other freely available design decisions (Wicherts et al., 2016).

The greater the use of this flexibility in analysis choices, the greater the risk of finding what null hypothesis testing calls a false positive result, or type-1 error, where you detect an effect in the sample, when there is no true effect in the population (Lindsay, 2015; Simmons et al., 2011). This use of flexibility to "fish" for significant results, is also often referred to as p-hacking (Harvey & Liu, 2021). However, Simmons et al. (2011) suggests that this exploratory behavior is not a product of malicious intent, but rather as a byproduct of ambiguity in decision making when designing a study, and the inherent desire of researchers to find a statistically significant result (Simmons et al., 2011).

Furthermore, Simmons et al. (2011) suggests that many commonly used practices can lead to false positive results, often without researchers necessarily being aware that their decisions have such prominent effect on false positive rates (John et al., 2012). Examples include the absence of a stopping rule prior to data collection, not disclosing all variables, eliminated observations, experimental conditions (especially failed manipulations), or failure to report statistics both with and without covariates. Combining flexible analysis choices like the ones mentioned above inflates the false positive rates in accord with how many researcher degrees of freedom was used (Simmons et al., 2011).

Multiple hypothesis testing

It is common practice in psychological research to apply null hypothesis significance testing to comparing means or other statistics between groups. This is done by testing the hypothesis that there is no difference between the groups (H_0), versus the alternative hypothesis that there is a difference between the groups. To do this, a test statistic is calculated, and the corresponding p-value is reported as evidence of significant differences,

resulting in a decision to discard or keep the null hypothesis depending on the magnitude of the p-value. The alpha criterion (the theoretical probability of making a false positive one is willing to accept) is commonly set to 5%, and the p-value corresponds to the probability to observe a difference that is equally large or larger than the actually observed difference (represented in the test statistic), assuming the null hypothesis is true (Curtiss, 1940). A common problem with null hypothesis testing is it is often used to answer the question "given this data, what is the probability H_0 is true?", while in reality it can only answer the question "Given that H0 is true, what is the probability of these (or more extreme) data" (Cohen, 1994).

Consequently, p-values where the null hypothesis is not discarded, yields little to no information, and should therefore be interpreted with care (Open Science Collaboration, 2015). Additionally, it is suggested p-values should be reported together with effect sizes and confidence intervals, which provides a better representation effect in the observed sample (Open Science Collaboration, 2015; Simmons et al., 2011).

To avoid problems when interpreting evidence in situations where the null hypothesis is not discarded, there are many ways to tweak the analysis, which increases the false positive rate, to find a significant difference in a p-value ≤ 0.05 (Simmons et al., 2011). As an example, by collecting and testing more than one outcome variable, assuming the outcome variables are reasonably correlated (0.5), the chance of finding at least one significant result can be doubled (Simmons et al., 2011). This is commonly known as the family-wise error rate used in multiple hypothesis testing, where the problem is usually circumvented by using certain statistical corrections such as the Bonferroni method (Savin, 1980).

The problem occurs whenever multiple hypotheses are tested, for example via collecting multiple outcome variables and/or running multiple tests for multiple predictor

variables and it is not always obvious when multiple hypothesis test correction is necessary (Simmons et al., 2011). As an example, this can be done by collecting two different outcome variables (such as collecting multiple different reports of MW, test them all, and only report significant effects). Another choice is to run multiple conditions, test them against each other and only report significant effects. In brain stimulation research, it is not uncommon to have multiple stimulation polarities, or even multiple brain areas (e.g., Kajimura et al., 2015, Filmer et al., 2021). By being able to test all the different conditions against each other (anodal vs baseline, cathodal vs baseline, anodal vs cathodal, DLPFC anodal vs inferior parietal lobule cathodal etc.) the chance to produce a false positive rate increases according to the number of conditions tested, unless proper corrections are applied (Simmons et al., 2011). However, testing participants in brain stimulation research is often time consuming and costly, therefore it is unlikely that sets of participants or conditions would be dropped. Instead, including multiple conditions rather comes at the cost of reducing the number of observations per cell, which can produce underpowered studies.

Statistical power and sample size

As mentioned, another type of analytical flexibility that can be employed, is interim testing (repeatedly conducting significance tests and stopping data-collection in case a significant result is obtained), which increases the chances of finding a false positive (Sanborn et al., 2014; Simmons et al., 2011). Until recent years, it has been uncommon that a fixed sample size or stopping rule was decided beforehand, and many researchers report that, wrongfully so, they think it has only minor implications for false positive rates (John et al., 2012). The stopping rule should be decided before data collection starts, independently of any data collected. By pausing often and repeatedly testing for significance the risk of finding a false positive due to random fluctuation increases. However, in some situations, it can be beneficial to employ an early stopping rule for ethical or financial reasons, to avoid collecting

an unnecessary large sample. This can be done by deciding on a stopping rule as part of a power analysis before data collection starts (for an example, see Boayue et al., 2019). Preferably, regardless of whether a data-dependent stopping rule is used, statistical power of a study should be calculated in any confirmatory study. Unfortunately, some researchers suggest that that it is likely that many researchers have a vague understanding of the concept of statistical power (Lindsay, 2015). The statistical power of a study is defined as the probability that a true effect of a certain size could be detected by a given statistical analysis should it exist in the sample. This is also referred to as the sensitivity (Cohen, 1992). To calculate statistical power, the researcher needs to have some knowledge about the population, such as an estimated effect size or variance that can be expected (usually from a pilot or previous study). Fortunately, steps have been taken recently by some top journals to require authors to provide evidence of statistical power, such as reporting a power analysis, before publishing (Lindsay, 2015).

In brain stimulation research, group sizes or observations per cell are usually on the lower end, and therefore also statistical power (e.g., Axelrod et al., 2015, 2018; Coulborn et al., 2020). This is very likely to introduce an overestimation of effect sizes when combined with flexible analysis methods, as smaller samples require larger effect sizes to reach the significance threshold (Boayue et al., 2019; Lindsay, 2015; Simmons et al., 2011). As an example, the original study by Axelrod et al. (2015) found an effect of $d = 1.24$ which is unusually large for psychological studies where one would typically expect to see effect sizes in the range of $d = [0.2, 0.6]$ (Open Science Collaboration, 2015). More specifically, in tDCS research, effects are usually small to medium with an average of $d = 0.4$ (Boayue et al., 2019; Horvath et al., 2015). When Axelrod and colleagues then moved on to replicate their own result, according to a power analysis based on their original effect size they increased they sample from 11 participants per group to 27 per group and found an effect size of $d = 0.97$

(Axelrod et al., 2018). However, power analysis based on inflated effect sizes will likely underestimate the required sample size and is a prime example of a negative cycle which contributes to why replication often fails to find the same effect sizes (Open Science Collaboration, 2015). Low powered research designs is suggested to be one of the main culprits in upwardly biasing effect sizes which are hard to replicate, supported by the fact that a large portion of replication research produce weaker effects than the original findings (Open Science Collaboration, 2015).

A good example of this is the study done by Boayue et al. (2019), which is a large-scale direct replication attempt of Axelrod et al. (2015). Since the original effect of d = 1.24 was assumed to be largely overestimated, Boayue et al. (2019) calculated their sample size to be $N = 96$ per group, which, according to their calculations, would be able to exclude $d = 0$ from the posterior high-density interval in the positive direction with a false positive rate of around 4% (Boayue et al., 2019). The effect subsequently found by Boayue et al. (2019) was $d = -0.11$, with the high-density interval not excluding zero. Additional Bayesian replication tests (Verhagen & Wagenmakers, 2014) suggested that it is 500 times more likely that the effect did not replicate, than that it did. While the original authors replicated their finding with an increased group size of 27 participants pr group to find an effect of $d = 1$, a more realistic effect of 0.4 (consistent with tDCS findings by Horwath et al., 2015), would require at least 78 participants pr group to find the effect with a power of 0.8 (one tailed t-test, alpha = 0.05; Csifcsak et al., 2019).

Publication bias and pressure to publish

The problems mentioned so far, are not unknown in psychological research, yet weak methodological or underpower studies are still being published frequently (Open Science Collaboration, 2015; Ferguson & Heene, 2012). Unfortunately, in the current publishing

model for scientific research, replicating existing research is not as desirable as publishing new and exciting findings (Ferguson & Heene, 2012; Simmons et al., 2011). The result is a publication bias towards positive findings and, therefore, significant results have a higher chance of being published than nonsignificant findings (Begg & Berlin, 1989; Braver et al., 2014). This publication bias is suggested to exist in editorial boards in prestigious journals, as well for individual researchers (Ferguson & Heene, 2012). Since replications do not add "new" knowledge, most top scientific journals are not as interested in replication articles compared to new and innovative research. New and innovative research have a higher chance of begin cited (which increases impact factor) and in turn increasing subscription rates and submission requests, creating a monetary incentive for journals to favor "new" and exciting research (Begg & Berlin, 1989; Callaham et al., 2002; Drotar, 2010). Additionally, researchers are under constant pressure to publish in order to advance their careers (Fanelli, 2010, 2011), and some researchers might simply just believe (not unreasonably so) that null results will not be published (Greenwald, 1975). Null findings in standard frequentist null hypothesis testing are also harder to interpret (Lindsay, 2015), and can therefore also be more lucrative to avoid, resulting in biasing published reports towards significant findings. In face of these strong incentives, some researchers may feel the pressure to modify their statistical analysis, fabricate data or selectively report variables to find significant results (p-hacking; Harvey & Liu, 2021).

Combining the pressure to publish on individual researchers, monetary incentives from journals to publish significant findings, the inherent explorative nature of finding significant results with a limited understanding of the association between the use of researcher degrees of freedom and the consequences it has on false positive rates, it is not surprising that a lot of psychological research fails to replicate. Thankfully, to ensure the good reputation of the scientific research process in psychology remains, certain steps have been

taken over the past decade to combat some of the problems with researcher degrees of freedom and low replicability.

Pre-Registration and Registered Reports

A new trend emerging in reporting psychological research, is to pre-register studies before data collection starts (Chambers, 2013; Pashler & Wagenmakers, 2012). Pre-registered studies aim at reducing the researcher degrees of freedom to counter inflated effects sizes and false positive results. By registering (and freezing) the study protocol, hypotheses, data collection and analysis plan before starting the study it is possible to reduce a lot of flexibility (Chambers, 2013). Additionally, some journals offer to publish in a format called Registered reports, where the study is registered, and peer reviewed before the start of the study. This is one way to ensure that a study will be accepted and published before starting, regardless of results, countering a publication bias towards significant results.

Registering the study forehand removes most experimenter and analysis bias, as some of the most important researcher degrees of freedom are taken away, such as flexibility in analysis and hypothesis planning, flexible stopping rules and selective reporting (Simmons et al., 2011). Additionally, due to the transparent nature of registration, the final report is also perceived as more credible (Simmons et al., 2011).

By combining the two concepts of registration and replication, a new type of report has quickly grown to be considered the best way of investigating novel findings in experimental psychology, pre-registered replications. Pre-registered direct replications are often considered the best way to confirm or deny the existence of preliminary effects and might be the best way to steer the field of psychological research out of the replicability crisis that is currently ongoing (Chambers, 2013).

In addition to direct replications where the goal is to keep everything as similar as possible to the original study, another popular way to replicate studies is a conceptual replication. A conceptual replication has much more freedom in how to design the study, where the main goal is to find the same effect, but different methodologies and experimental setup can be used, as well as the option to extend the scope of the theory rather than just reproducing the previous results (Simons, 2014). Conceptual replications are not as robust as direct replications when it comes to confirming or denying the existence of an effect, due to the fact that different methodologies can introduce added variance which can interfere with the reproducibility of the effect in question (Simons, 2014). However, conceptual replications often assume that the original effect is robust enough to appear with differences in study design. Otherwise, there is an argument to be made that there is little to no use for a fragile effect which only appears under perfect conditions in the lab.

Replication brain stimulation research

As an example of a conceptual replication, the lab of Boayue et al. (2019) went on to improve on some of the original methods of Axelrod et al. (2015), in an attempt to replicate the finding of increased MW caused by tDCS over the DLPFC. They improved the tDCS technique by creating a more focal and powerful stimulation protocol. They also improved the task by changing the SART for the FTRSGT, and to their surprise they found an effect of tDCS over the DLPFC modulating the propensity to MW (Boayue et al., 2020). This effect was however in the opposite direction of the original finding by Axelrod et al. (2015), namely that HD-tDCS over the DLPFC reduced the propensity to MW. While the methodology was quite advanced compared to other studies in the field, using a highly sensitive task and analysis method, substantial sample size (30 observations pr cell), double blinded design, within-subject repeated-measures design (comparison against individual baselines), as well as reducing their researcher degrees of freedom by establishing analytical choices based on pilot

experiments that did not include brain stimulation (Boayue et al., 2020). Unfortunately, the final analysis pipeline was not registered, and thus only provides preliminary evidence. If the arguments made so far in this thesis is to be believed, the only way to get closer to the truth is to let the effect of HD-tDCS over the DLPFC to influence MW weather the storm of a pre-registered direct replication attempt.

Current study

Considering the problems currently highlighted, and with healthy skepticism in mind, we decided that the only solution to properly investigate if HD-tDCS over the DLPFC can reduce the propensity to mind-wander is to conduct a pre-registered, high-powered direct replication of Boayue et al. (2020), which we did. The report was accepted by the journal NeuroImage: Reports, and below you fill find the methods and results of said registered report.

Methods

To adhere to Open Science principles, all the planned analysis, hypothesis, materials, and raw data is freely available in an open public repository at Open Science Framework (OSF) https://osf.io/cv24f/.This study and data collection is part of a bigger project set out to investigate the intricate neurophysiological relationship of executive functions and MW. To this extent electroencephalogram (EEG) and pupillometric data was also collected, but is outside the scope of this thesis, and therefore also not reported.

Participants

Participants were healthy adults aged 18-35, who were recruited randomly through social media and networking. Exclusion criteria were as follows: Psychiatric/neurological disorders currently or in the past (such as, bipolar disorder, epilepsy, severe head trauma,

migraine, depression, brain surgery etc.), using medication which affects the central nervous system (such as antidepressants or antiepileptic drugs), being under the influence of psychotropic drugs, with an exception made for caffeine and nicotine, not being fluent in either English or Norwegian. In addition, a few minor criteria must be met on the test day to be included and tested, the participants must have slept enough the night prior to testing, not be extremely hungry, not have too much eye makeup (for the eye tracker) and have good or corrected eyesight.

The first set of criteria regarding mental and neurological health was to ensure a homogenous sample of healthy adult brains to remove any unwanted uncontrollable variance within the sample. The second set of criteria for the test day were included to ensure everyone understood the task equally (language proficiency and eyesight), as well as in somewhat the same mental state when it comes to primitive needs such as hunger and sleep. Sleep and wakefulness are crucial when it comes to investigating task performance, attention and mind wandering (Jubera-Garcia et al., 2021) and while participants were testing mostly during the normal the working hours (9-17), studies have shown that performance can vary depending on the time of day and our circadian rhythm (Blatter & Cajochen, 2007; Peñalva et al., 2003; Åkerstedt, 2007). To make sure this was kept as controlled as possible, participants were asked if they had slept enough and felt awake/rested before the task started and were allowed to take breaks during the task but between the blocks (baseline/stimulation), however no participant opted to take breaks longer than 15 minutes. Any participant who closed their eyes and failed to follow instructions (tapping their fingers to the sound of the metronome) were deemed sleeping, paid for participating and their data excluded and replaced by a new participant immediately.

Design and procedure

Lab, location, personnel and procedure

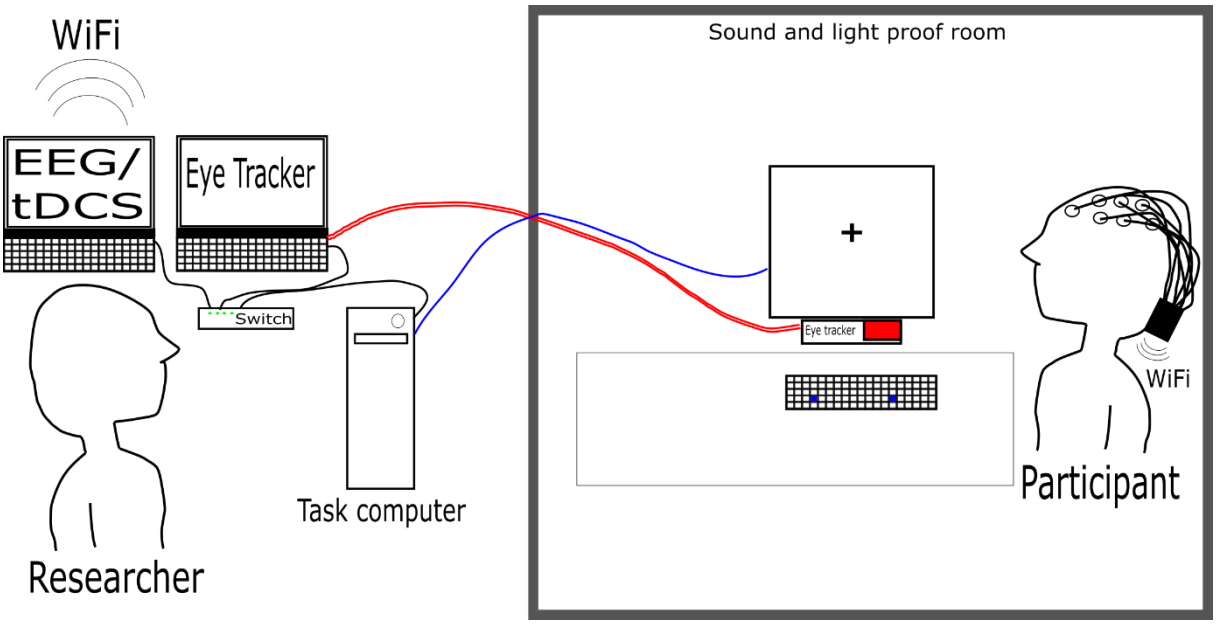
The experiment was conducted at The Arctic University of Norway (UiT), in an isolated lab protected from light and sound, providing a non-disturbing environment for testing (see figure 1), the exact same lab and environment as Boayue et al. (2020). During testing the outer lab door was locked and an occupied light was lit to ensure the experiment could not be interrupted. Outside of testing the lab always remained locked. During testing the participant was fitted with the electrode cap and electrodes while they read instructions, before they completed a quick training session (1 minute) and answered a mini quiz to ensure they understood the task. Once completed the participant could start with the baseline block. To start the experiment the participant was put in the testing room in front of a 19" flat screen monitor at a fixed 70cm distance (with the possibility to adjust height with an adjustable table), fit with a high-quality stereo headset (Multi-Function Headset210, Trust International B.V., Dordrecht, Netherlands), and before the experiment started, the experimenter stepped outside. The experimental task and stimulation protocol was started remotely from outside the room, and the instructions were presented again on the task computer in written text before each block (baseline, stimulation). The experiment computer was set up disconnected from the internet and free from any disturbing background processes. The experimental task was run in psychoPy3 (release v2020.1.3).

Data was collected by four trained experimenters, and all experimenters received standardized instructions and were required to practice on at least two pilot subjects before collecting real data. All participants received instructions in a written format (either Norwegian or English) to keep experimenter influences at a minimum, however participants could ask for clarification at any point during before the baseline block and in the break between the two parts. Additionally, the experiment was running during the ongoing Covid-19 pandemic, which required contagion preventive measures such as extensive cleaning, and

wearing extra protective equipment (gloves, mask, coats). Approval for these contagion preventive measures were obtained from the Faculty of Health Sciences at UiT The Arctic University of Norway, ahead of data collection.

Figure 1

Illustration of the lab setup and apparatus



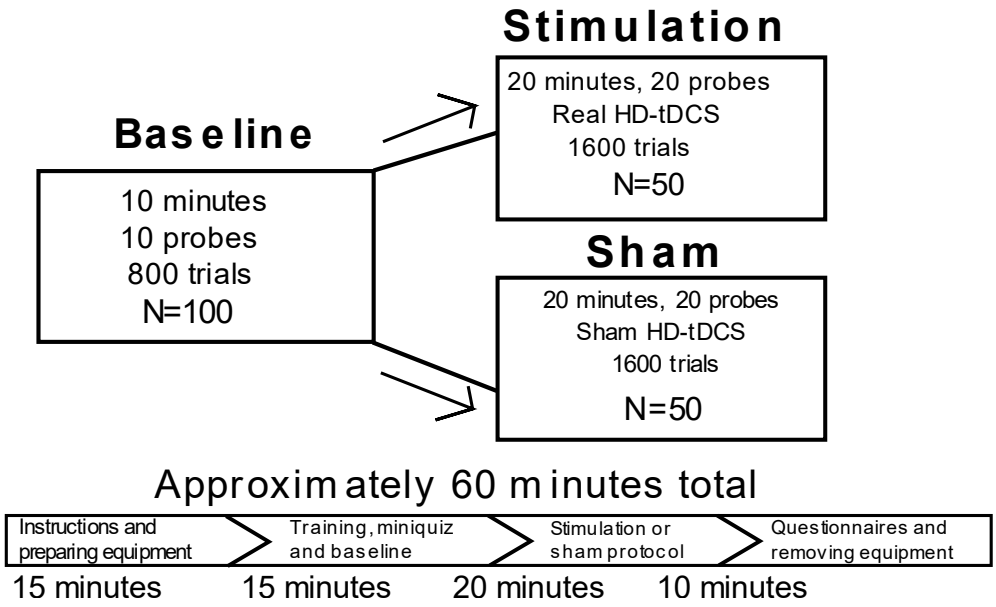
Note. The participants were sitting in an isolated chamber while the researcher was monitoring progress outside. The experiment was part of a larger project, therefore EEG/Pupil data was also collected.

Study design

The experiment was set up as a mixed design with one between-subject factor (Condition: sham vs. real stimulation) and one within-subject factor (Block: baseline, stimulation), see Figure 2.

Figure 2

Flowchart of the experimental session



Note. All participants started by completing a training session and a mini quiz before starting the baseline part.

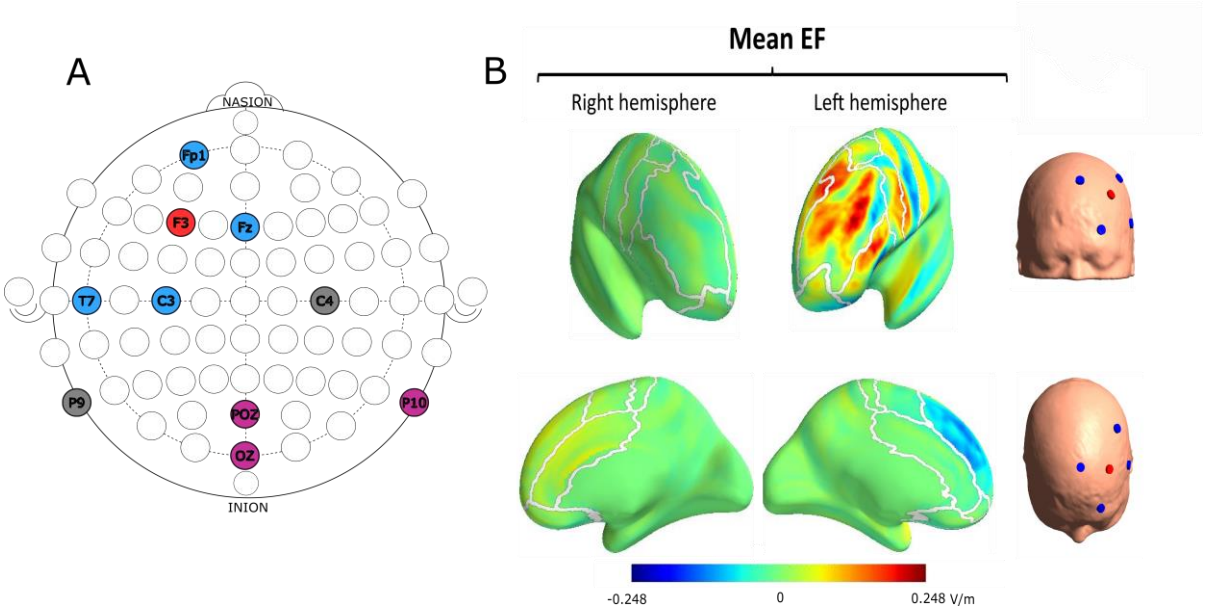
Blinding

The experiment was conducted in a triple-blind manner; both experimenter and participant were blinded to the experimental condition, and additionally the main analyst (MM) was blinded to the experimental condition by recoding the conditions to condition A and B by a third investigator (GC). Only after the main analysis was complete, were the groups revealed. To successfully blind the participants a mild anesthetic cream (EMLA) was applied directly under the stimulation electrodes, as side effects such as irritation or itching has been shown to compromise the blinding (Turi et al., 2019). To test the efficacy of the blinding procedure participants were asked to report whether they think they received real stimulation or not on a 7-point Likert scale ranging from 1 which was “definitely not stimulation” to 7 which was “definitely stimulation”.

HD-tDCS protocol

Figure 3

Positioning of electrodes (A) and electric field produced by the HD-tDCS (B)



Note. A) Electrode positioning for the HD-tDCS: Anode F3, Cathode T7, C3, Fz, Fp1, reference electrode above the mastoid at P9, and ground on C4. B) Stimulated electric field coverage reused with permission from Csifcsák et al. (2018), realistic head models created from MR images of healthy adults. The normal component of the electrical field is perpendicular to the cortical surface and represents the current flowing inwards or outwards from the gray matter. Due to the cytoarchitecture of the cerebral cortex, these currents are primarily associated with affecting the excitability in pyramidal cells (Rahman et al., 2013).

HD-tDCS was delivered with a 4x1 montage (see figure 3), using a Starstim Neckbox (Starstim tDCS, NE Neuroelectrics) and PISTIM EEG & tDCS Ag/Ag/AgCl electrodes (12mm diameter). The HD-tDCS montage has 1 anode surrounded by 4 cathodes, such that the electrical current will enter through the anode, distribute across the scalp and brain tissue, before returning through the cathodes. This setup creates a strong electric field around the anode and the surrounding brain region (Csifcsák et al., 2018). The real stimulation condition was set to stimulate continuous stimulation at 2mA intensity for 20 minutes, plus an additional 30s ramp up and 30s fade out period (total 21 minutes). The sham stimulation

condition received the same ramp up period, but then produced no electrical current for the remainder of the block.

The electrode position was fitted with a position cap, in accordance with the international 10/20 EEG system. Head circumference was measured to decide cap size, with standardized size cutoffs (54cm = M, 57cm = L, 62+cm = XL). When the correct cap was fitted on the participants head, symmetry was checked by measuring distance between nasion, inion, preauricular point and respective electrode positions (Fz, Cz and Oz). Instruction on how to measure head size and fit the positioning cap was standardized and practiced by all experimenters and can be found on OSF (https://osf.io/u8n7x/).

The cognitive task

The task is finger-tapping random sequence generation task (FT-RSGT), which is a remake of two classical tasks, the random number generation task (Baddeley et al., 1998; Towse, 1998) and a finger-tapping task (Kucyi et al., 2017; Seli, Cheyne, et al., 2013). The task requires participants to press one of two buttons with their index fingers in response to an ongoing rhythmic metronome, as precise as possible, while the resulting pattern of finger taps is kept as random as possible. The task it easy to understand, yet monotonous over longer periods of time creating the opportunity to MW. Additionally, creating random patterns draws heavily on executive resources (Teasdale et al., 1995). In turn we assume that the randomness in the pattern created is related to how much executive resources was used on the task. Studies have found that random sequences created during attentional lapses often are a result of autopilot behavior which typically is less random (Boayue et al., 2020; Teasdale et al., 1995). In addition to measuring randomness, the during the FT-RSGT it is also possible to calculate the behavioral variability (BV) related to the responses to the target stimuli (metronome sound). Previous research has found that when attention is diverted away from the task, the

BV increases accordingly (Kucyi et al., 2017; Seli et al., 2013). Due to the fast-paced nature of the task, the FTRSGT provides measurements of high temporal resolution, making it easy to investigate fluctuations in executive resources and behavioral variability which in turn can be correlated to MW.

To supplement the measurements provided on the behavioral level during the FT-RSGT, experience sampling though probes were also dispersed throughout the task. These probes appear at random or fixed intervals during the task asking the participants to answer a question such as “Where was your thoughts directed right before this question appeared?” With the responses available being a rated on a Likert scale ranging from 1 “completely on task”, to 4 “completely off task”. This provides an extra subjective measurement by asking participant to introspectively report their attentional state. In the present study, these probes appeared at a random interval between 40-80 seconds (uniformly selected).

Measured variables

The main dependent variable we collected is the answer to the intermittent thought probes. In total the participants replied to 30 probes split into 10/20 probes between the baseline and the stimulation block respectively.

Behavioral Variability

Behavioral variability was measured as the standard deviation of the inter-tap-intervals (ITI) using the last 25 targets preceding a thought probe. No responses or trials were excluded during the calculation of the BV measure. That way, both missing responses as well as double-taps acted towards increasing the measure. This procedure was identical to the one used in Boayue et al. (2020).

Approximate Entropy

To measure how random the participants created their sequences, a statistic called approximate entropy was used (AE; Pincus, 1991). Approximate entropy can be used to evaluate the irregularity in a given sequence and is parameterized by the parameter m. AE(m) then measures the logarithmic frequency with which blocks (with length m) that are close together, remain close together for blocks augmented by one position (Pincus & Kalman, 1997; Pincus & Singer, 1996). In turn we get a measurement AE(m) where higher values represent more irregular sequences, providing a measure of randomness. The parameter m (block length) has been optimized by Boayue et al. (2020), which determined the optimal length to be m=2, which also was used for this study. AE was calculated using subsequence length m=2, on the 25 preceding taps before each probe and is transformed using the formula $\mathrm{AE}_{trans} = -\log\left(log(2) - AE_{raw}\right)$.

Deviations from the original study

While the study was designed to be a direct replication of Boayue et al. (2020), certain subtle differences were unavoidable. Due to the nature of the multimodal data collection, participants were asked to keep their chin on a headrest for the Eye tracker during the experiment, and the eye tracker were calibrated before the onset of the first block, which they did not have to do for the original study. Additionally, to record EEG data, the participants had to perform the baseline block with the electrode cap on, which differs from the original study where the cap was only worn during the stimulation block. Since the cap was fitted between the baseline and stimulation block in the original study, there might be a slight time difference (1-3 minutes) between the onset of the stimulation block after the end of the baseline block. Participants were allowed breaks in-between blocks, meaning there was not a standardized time between the blocks in the first place, and neither do we have any reason to assume this time difference is important.

Another difference is that due to the covid-19 pandemic which was at its peak during the time required both participant and experimenter to wear protective equipment (mask, gloves, coats), however the participants were allowed to remove the mask during the cognitive task. Regardless of these subtle differences, neither of them should compromise the replication attempt and any potential effect caused by the HD-tDCS should be robust enough to withstand these small differences in experimental protocol.

Statistical methods

Four hypotheses were pre-registered; (1) we expected mind wandering to be reduced in the real vs sham stimulation condition, (2) we expected behavioral variability (BV) to be increased preceding mind-wandering episodes, (3) we expected randomness in the sequences participants create (operationalized by AE) to be reduced preceding mind-wandering episodes and finally (4) we expected to observe a positive interaction effect between BV and AE. To test these four hypotheses, we used a Bayesian hierarchical ordered-probit regression model with random intercepts and experimental block nested within subject (see Boayue et al., 2020). This model included behavioral variability (BV), Approximate entropy (AE), their interaction, Trial (probe number), Condition (real vs sham stimulation), Block (baseline vs stimulation) and the interaction between Condition and Part. From this model we extracted posterior mean and highest density intervals (HDI) of the regression coefficients as well as evidence ratios (ER). The model was implemented in R programming language (R Core Team, 2015) and Stan (Carpenter et al., 2017), using the packages BayesFactor (Morey & Rouder, 2015), rstan (Stan Development Team, 2016) and brms (Bürkner, 2017). Assessment of blinding was done in JASP (JASP Team, 2021).

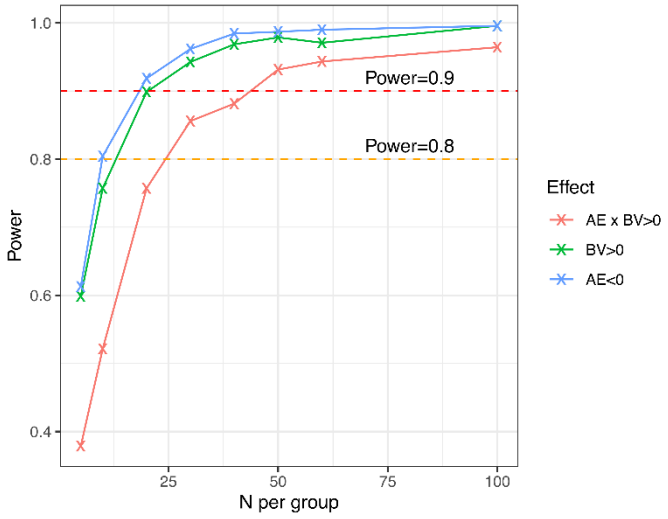
Statistical power

To estimate the power of the study, the posterior distribution from the model calculated by Boayue et al. (2020) was used to simulate random datasets, to then fit back the analytical model and estimating the probability of finding an effect where 95% of the posterior density is in the expected direction. Full details of the power analysis can be found at OSF (https://osf.io/3fcsx/). For practical reasons (time-constraints due to study programs, and uncertainty due to the ongoing pandemic) we decided to set a maximum sample size of $N = 100$ (50 per condition), resulting in a total of 100 valid datasets.

For the effects for BV, AE and their interaction random datasets were created by drawing samples of increasing sample-sizes from the posterior distribution from the previous study and creating though-probe responses using the model predictions (figure 4, see https://osf.io/3fcsx/ for details).

Figure 4

Power-curves for the effects of BV, AE and BV x AE calculated based on the full posterior distribution.

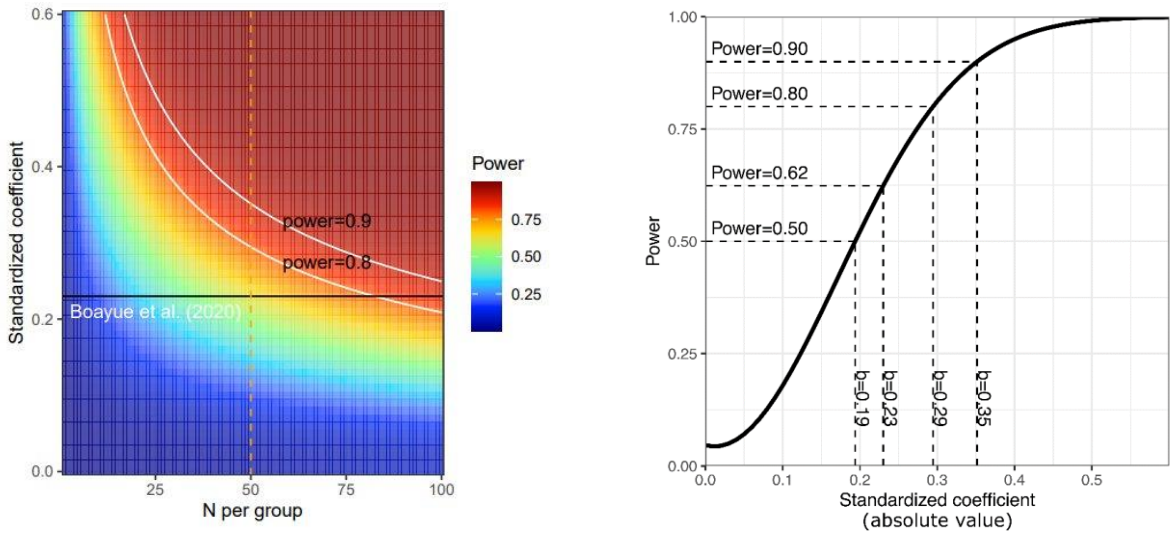


To calculate the power required for the effect of HD-tDCS on MW, we ran power-calculations for a grid of sample-sizes and effect-sizes for one week on a server with 80 cores

and used a parametric model fit to provide an approximation to the actual power-curves. Based on this approximate power-analysis we calculated the power-surface as a function of sample- and effect-size (figure 5).

Figure 5

Power-surface as a function of sample and effect size (left), power of the effect of HD-tDCS on MW for $N = 50$ *(right)*



Since we set a maximum of $N = 50$ participants pr condition for practical reasons we investigated power for $N = 50$ in more detail by plotting power as a function of standardized regression coefficients (figure 5). We have approximately 62% power to detect a real effect-size as large ($b = -0.23$) as in Boayue et al., (2020). The chance to detect effect sizes as small as $b < -.19$ is around 50%. To achieve 80 and 90% power, the real effect-size would have to be $b = -.29$ and $b = -.35$, respectively. Since the sample-size required to achieve 80% power in all our target analyses exceeds our pre-specified $N_{max} = 50$ per condition, we stand by our decision to collect data until 100 valid datasets ($N = 50$ per condition) have been accumulated.

Results

Demographics

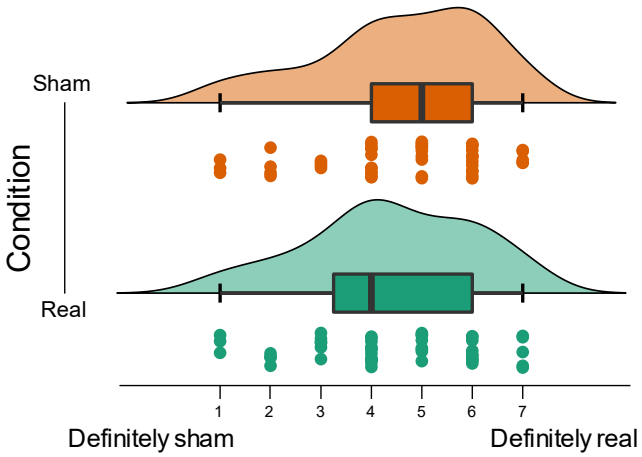
The final sample consisted of 100 participants (female = 54), aged 18-36 (mean = 24.58). Due to the exclusion criteria no datasets were excluded, however in total eight datasets were replaced in total (7 immediately during data collection, and 1 after data inspection). Two participants were replaced due to failure to provide a complete dataset (lack of responses to target stimuli), four participants were accidently tested with the wrong test protocol or wrong protocol sequence, and one participant was familiar with the study and study hypotheses. Additionally, there were four problematic datasets (only one type of response, always responding to be on task and double the number of expected responses), however they did not break any exclusion criteria and were therefore included in the final sample. One additional dataset was excluded (and replaced) after data inspection due to lack of responses to target stimuli. All datasets, including the excluded ones can be found on OSF (https://osf.io/cv7qm/).

Blinding

To assess the efficacy of the blinding we tested the hypothesis that the participants in the real stimulation condition scored higher on the 7-point Likert scale compared to the participants in the sham condition. To test the hypothesis, we ran a two-sided Bayesian Mann-Whitney test which supported the null hypothesis (BF_{01} = 4.16, see figure 6), providing evidence that the blinding was successful.

Figure 6

Distribution of the replies to the blinding questionnaire



Note. Boxplots and density curves modeled as replies to the Likert scale asking whether the participants thought they received real stimulation or sham stimulation after the experiment.

Effect of HD-tDCS

In line with the registered analysis plan, we used a hierarchical ordinal regression model with AE, BV, their interaction, Block (baseline vs stimulation), Trial (probe number), Condition (sham vs real stimulation) and their interaction to analyze the responses to the thought probes. The model yielded an adequate fit to the data, R^2_{bayes}= 0.35.

There was no difference between the two conditions in the baseline block in regard to self-reported MW (b = -0.05 [-0.35, 0.23], ER_- = 1.54). We did not find any evidence of an effect of HD-tDCS on the propensity to reduce mind wandering in the stimulation block (*b* = 0.00 [-0.19, 0.19], ER_- = 1.01). We confirmed the hypothesis that BV was increased preceding self-reporting MW episodes compared to on-task periods (*b* = 0.19 [0.14, 0.24] $ER_+ = \infty$), and that AE (utilization of executive resources) was reduced prior to self-reported MW episodes (*b* = -0.07 [-0.11, -0.03], ER_- = 362.64). We did not observe any evidence of the expected positive interaction effect between BV and AE (*b* = 0.00 [-0.03, 0.04] ER_+= 1.36).

We also observed strong evidence that MW increased as the task duration increased, compared with baseline, we observed increased MW in the stimulation Block (*b* = 0.27 [0.13,

Note: calculating the R^2 were done using the "bayes_R2" function in the brms package in R, where predictors are treated as continuous variables.

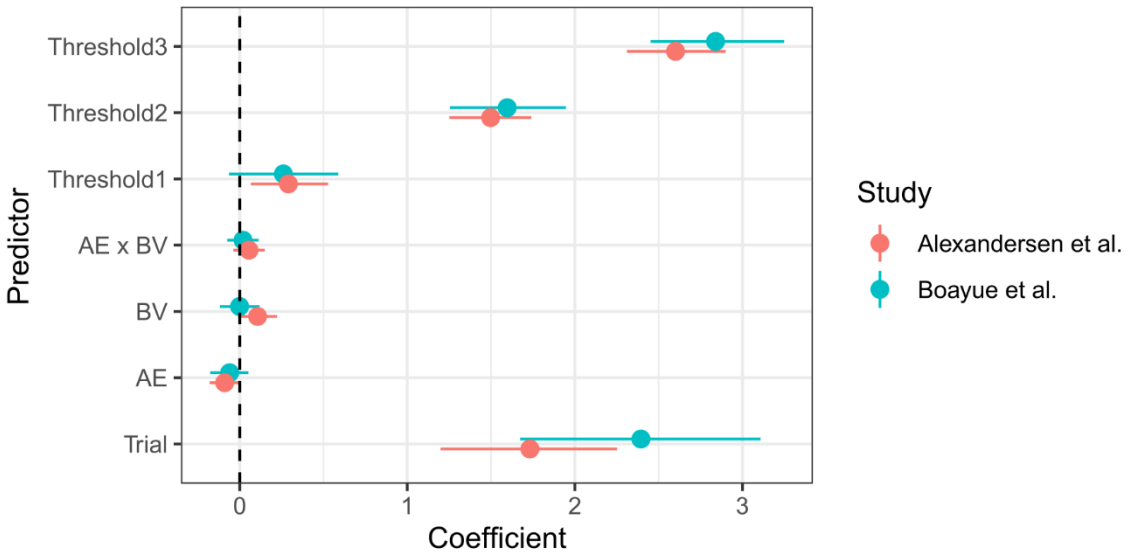
0.40], $ER_+ = 1332.33$), and while this was expected due to the increased duration of the stimulation part, we also found a positive effect of Trial within-blocks (*b* = 0.81 [0.68, 0.95] $ER_+ = \infty$).

Comparing of the sample with Boayue et al. (2020)

To assess whether the sample in our study was homogenous to the sample in Boayue et al. (2020), we compared the two baseline Blocks in both studies in a new hierarchical ordered probit model with probe response as dependent variable, and BV, AE, Trial and Experiment as predictor variables (figure 7). The Experiment variable was allowed to interact with all the other predictors to check if they had different relationships with MW across the two experiments. Out of all the predictors only the BV variable barely excluded zero (BV: *b* = -0.10 [-0.18, -0.02]), the other coefficients did not (AE: *b* = -0.05 [-0.13, -0.04], Trial: *b* = 0.03 [-0.23, 0.28], BV x AE: *b* = 0.03 [-0.04, 0.10]). We decided the two samples were comparable and combined them to reach a full sample of N = 160 and re-run the original model with all the predictors and allowed all the predictors except Block and Condition to interact with the experiment variable. The new full model showed that the 95% HDI on the effect of HD-tDCS on MW did not exclude zero (*b* = -0.08 [-0.22, 0.06] $ER_- = 4.69$).

Figure 7

The predictor coefficients from the baseline block in the two studies



Note. Threshold 1-3 is the different thresholds that differentiate the probe caught ordinal responses of the Likert scale in the hierarchical ordered probit model.

Discussion

The main goal of the present study was to replicate the effect found by Boayue et al. (2020), that HD-tDCS over the left DLPFC can reduce the propensity to MW, with a high-powered pre-registered study protocol. We did not observe any evidence that HD-tDCS can reduce or influence the propensity to MW. Additionally, when the samples of the original study and the present study were combined (to reach a full sample of $N = 160$), the original effect of HD-tDCS reducing propensity to MW found by Boayue et al. (2020) disappeared. This reinforces the claim that any positive effect of tDCS on mind-wandering, and cognitive functions in general should only be accepted after a registered direct replication, which currently seems to be lacking. We are therefore forced to conclude that the earlier finding in Boayue et al. (2020), despite the sophisticated methodology was highly likely to be a false positive effect. These findings serve as a reminder that even rigorous methods can produce false positive results with only a small amount of researcher degrees of freedom.

Failing to replicate effects of tDCS on mind wandering

As mentioned in the introduction, the best way to confirm or deny the existence of an effect is by a rigorous high powered registered direct replication. Do the results above then mean that tDCS cannot reduce MW? Not necessarily. Instead, we shall highlight possible problems, and consider alternative explanations as to why effects found in brain stimulation and tDCS research may fail to replicate.

Brain stimulation montage and efficacy of tDCS

The first conclusion is to assume that tDCS is not powerful or specific enough to modulate brain activity. Indeed, problems with the traditional bipolar tDCS montages have been highlighted, such as low focality at the cost of stimulation intensity (Csifcsák et al., 2018; Datta et al., 2008, 2009; Nathan et al., 1993). However, before we write that off that as the sole reason, we should entertain the possibility that the current experimental paradigm is not perfect and can still be improved upon. The methodology of tDCS have improved in recent years, and more researchers are moving away from the traditional square "pad" montage, to more focal stimulation paradigms such as the "ring" configuration of HD-tDCS such as the 4x1 montage used in the current study (Csifcsák et al., 2018; Datta et al., 2008, 2009; Nathan et al., 1993). While increased focality is a positive development when targeting specific brain regions, recent research suggests it might also be counterproductive unless done correctly, as increased focality relies on precision targeting of cortical areas (Trembley et al., 2014). Due to the large variances in between-subject variability in brain anatomy, increasing focality while using the same scalp positions across subjects might end up targeting different cortical areas altogether (Tremblay et al., 2014). To solve this problem, some research has suggested to use prospective electric field modelling for individual participants (Csifcsák et al., 2018; Evans et al., 2020; Rasmussen et al., 2021). Prospective electric field modelling is a technique that allows researchers to model the spatial distribution of the electrical field induced by the stimulation for each individual participant ahead of actual stimulation. This

technique works by incorporating brain scans (such as fMRI) to map out brain structure, together with knowledge of the conductivity of the individual layers of the head and brain (scalp, skull, white/gray matter, and cerebrospinal fluid) to estimate how an electric field would distribute across the cortex (Klooster et al., 2016). Prospective electrical field modeling can then be used to individually place electrodes for each participant, to make sure the same underlying cortical areas are targeted.

While prospective electrical field modeling was not used in the present study, all participants heads were measured, a cap was fitted (small, medium or large) and carefully position according to symmetry before the electrodes were mounted to ensure electrodes were placed directly above the targeted brain region. While there might have been some discrepancy between participants head sizes and how well the electrodes were fitting, this difference should not have been large enough overpower the entire effect of HD-tDCS. Nevertheless, prospective field modeling is promising when it comes to reducing some of the between subject variability by ensuring that the correct brain regions are targeted across participants, instead of relying solely on the international 10/20 system to position electrodes. Increased focality of stimulation montages, together with increased precision from prospective field modeling could be the answer to some of the problems that tDCS suffers from.

In the end, it is hard to get around the fact that tDCS might not be best suited to modulate MW and is often favored for it’s easy to use setup and portability (Hordacre, 2018; Simon & Bikson, 2019). Other NIBS techniques such as transcranial magnetic stimulation (TMS) have a larger body of evidence supporting its efficacy on modulating brain activity (Cole et al., 2015; Loo & Mitchell, 2005; Zrenner et al., 2018), although not currently in MW research. While TMS is not as portable as tDCS and takes longer to configure, some forms of TMS was approved by the US food and drug administration in 2014 for use by practicians in

therapeutic use against mood disorders such as depression (Cole et al., 2015; Levkovitz et al., 2015). Perhaps one option is to consider switching stimulation protocol altogether if the goal is to change the propensity to MW with non-invasive brain stimulation. Alternatively, there are other methodological options to consider first, before writing off the effects of HD-tDCS on MW completely, such as improving the statistical methods.

Statistical methods and statistical power

Another problem with comparing and replicating brain stimulation research, is that there is also a large variability in which statistical methods are used, as well as that a lot of brain stimulation studies are underpowered. Dr. Stephen Lindsay, editor of *Psychological Science* named a "troubling trio" of replication problems: Low statistical power, surprising results, and p-value only slightly under 0.05. There seems to be an increase regarding the appreciation of the importance of sufficient statistical power, as researchers are getting more familiar with the concept and more journals are starting to list evidence of statistical power as a requirement to publish (Lindsay, 2015). When it comes to problems regarding null hypothesis testing and interpretation of p-values, a promising alternative is a Bayesian approach to statistics.

Unlike null hypothesis testing, Bayesian statistics does not rely on asymptotics, and is therefore known to yield more robust results in studies with relatively small samples (McNeish, 2016; Stegmueller, 2013). However, Bayesian statistics uses a prior to update the posterior belief, and in small samples the estimates will be sensitive to the strength of the prior (McNeish, 2016). One of the largest criticisms of Bayesian statistics, is the subjectivity introduced when selecting a prior, and therefore Bayesian approaches to solve issues of small samples should be used with care (McNeish, 2016). To circumvent this problem, researchers have suggested to use weakly informative priors to solve sparse data problems (Hamra et al.,

2013). Additionally, since the prior used in Bayesian statistics often is based on the results from previous findings, Bayesian statistics a good fit for replication research, and special Bayesian replication tests have been developed for this purpose (Verhagen & Wagenmakers, 2014).

When it comes to the problem of interpreting null results in null hypothesis significance testing (e.g., Cohen, 1994; Ferguson, 2009; Lindsay, 2015), Bayesian statistics offers a solution to that problem. In Bayesian statistics, instead of giving the likelihood that the data being tested is extracted from a sample where the null hypothesis is true, Bayesian hypothesis testing instead gives the probability that the test results support the null hypothesis as opposed to the alternative hypothesis (Masson, 2011; Wagenmakers et al., 2011). An increased use of Bayesian statistics instead of null hypothesis significance testing might therefore be able to reduce the current aversion to report evidence in favor of the null hypothesis which exists in the literature (Ferguson & Heene, 2012), and consequently reduce publication bias.

Some researchers argue that an increased use of Bayesian statistics introduces more researcher degrees of freedom by giving a selection of which prior to use and providing another set of analysis (to potentially selectively report, Simmons et al., 2011). However, this problem is easily circumvented by registering the analysis plan beforehand, or by reporting both (and preferably all) analyses done. An even better option is to run a sensitivity analysis (or a comprehensive specification curve analysis, see below) for the prior selection, or for analytical choices made in frequentist models to check the robustness of the conclusion (Simmons et al., 2011; Simonsohn et al., 2019). Considering all the flexibility in arbitrary choices made when setting up an analysis, it is important to check whether the conclusions drawn are a direct result of the analytical choices made, or if the conclusions are robust enough to transcend these analytical choices. A good way to test the robustness the effect in

question, and in turn the conclusions drawn, is specification curve analysis (Simonsohn et al., 2019). Specification curve analysis is a way to test analytical choices made such as transformations, choice of covariates, interactions, dropping of outliers, or for Bayesian statistics the range of the prior, to check if changing arbitrary analytic choices changes the conclusion of the analysis (see Simonsohn et al. (2019) for an example). However, specification curve analysis introduces a set of technical problems and increases the complexity of the analysis, in a field which is only recently starting to familiarize with the concept of statistical power and researcher degrees of freedom. Nevertheless, specification curve analysis can be a great alternative for studies that are unable to pre-register their analysis pipeline to support the robustness of their findings. In the current study, the full analysis pipeline was registered, and peer reviewed before any data was looked at, and therefore removed any flexibility in analytical choice, and in turn ensuring that the false positive rates and false negative rates would be exactly the values set in the power analysis.

Statistical modeling of ordinal data

Another methodological concern that can cause problems is when ordinal data is treated as metric, as highlighted by Liddel & Krushke, (2018). As most brain stimulation and MW research relies on ordinal data as their main outcome variable in form of thought probe responses, this can cause problems which leads to erroneous statistical modeling. To avoid using ordinal data in metric models (which assumes data is continuous and on interval or scale level), a commonly used practice is to average ordinal variables to create a continuous metric variable that can be used in regular regression models. While this seemingly seems to circumvent the problem, averaging ordinal data to mask it as metric is prone to erroneous results (Liddell & Kruschke, 2018). Firstly, averaging causes unfavorable data loss, such as time on task effects (e.g., where performance decays over the duration of the task), secondly the underlying ordinal data does not have equal distance between levels, a true zero and is

often not normally distributed which are all requirement for standard regression models. Violating such assumptions can easily lead to false positives (type-1 errors) or false negatives (type-2 errors) when reporting results (for an overview, see Lidell & Kruschke, 2018). Instead, it is suggested that researchers use updated statistical models (e.g., the hierarchical ordered-probit regression model; Boayue et al., 2020) to circumvent errors caused by using ordinal data as metric.

Blinding

Another concern which is heavily debated in NIBS research is blinding. Poorly blinded studies may suffer from both intentional and unintentional bias from both experimenter and participants, and thus also influence the replicability and validity of NIBS research. Even though newer tDCS stimulators have the option to operate in double blind mode, studies have found that discomfort during stimulation and perceptual differences (such as redness of the skin) may compromise blinding (Turi et al., 2019). NIBS studies with compromised blinding can be problematic, as an example, a study was recently published evidence suggesting that subjective prediction of tDCS was a better predictor of MW than actual tDCS (Fassi & Kadosh, 2020). Normally in NIBS research, blinding is assessed by asking the participants to introspectively assess whether they believe they received stimulation or not. Instead, Fassi & Kadosh (2020) suggested introducing subjective beliefs about the intervention as a factor in the main analysis. By reanalyzing an open access dataset published by Filmer et al. (2019), they found that participants subjective belief about receiving stimulation affected their performance more than the actual stimulation they received (Fassi & Kadosh, 2020). To support their claim that blinding might have been compromised, Fassi & Kadosh (2020) also argued that in the original study by Filmer et al. (2019) the participants received tDCS stimulation offline before the task, which could have

given participants opportunity to fully focus on detecting the sensation of the stimulation, thus compromising the blinding.

However, the claim made by Fassi & Kadosh (2020) was later criticized (Gordon et al., 2021). The main criticism was that the original study by Filmer et al. (2019) used five different stimulation protocols, and Fassi & Kadosh (2020) did not assess the key contrast between the main condition and sham, instead they examined all conditions together in their analysis. Gordon et al. (2021) then went on to replicate the original finding from Filmer et al. (2019). They found that active intervention was a better predictor of stimulation effects on MW compared to subjective beliefs about the intervention, even when using the approach suggested by Fassi & Kadosh (2020). However, they concluded that it is important to control for and understand the possible effects of subjective belief in sham-controlled studies (such as the present study). Furthermore they suggest that the best way to prevent issues with subjective beliefs and problems with blinding, is the inclusion of active control conditions (Gordon et al., 2021).

While the present study did not include an active control condition, all participants were made oblivious to any sensation associated with active stimulation by applying a local anesthetic directly under the electrodes 20 minutes before stimulation. Additionally, even though subjective beliefs about intervention was not included in the main analysis, blinding efficacy was checked in a separate analysis, which clearly showed evidence of effective blinding. In fact, participants in the sham group reported higher scores (reflecting belief that they received real stimulation) to the blinding questionnaire compared to the participants who received real stimulation, though there was no significant difference between the groups. Therefore, we conclude that since the participants failed to recognize which group they were assigned to, the blinding was successful.

Self-reported data

Brain stimulation and MW research suffers from the subjectivity and uncertainty in self-reported data. Due to the subjective nature of the experience of MW, common practice to study MW is to rely on self-reported measurements to quantify MW (Seli et al., 2013; Smallwood & Schooler, 2006), this study is no exception to this, as the main regression model is reliant on self-reported thought probes as the main outcome variable. It has been suggested that it might be time for MW research to use machine-learning approaches and rely on behavioral or neurophysiological measures to circumvent this problem and get rid of though probes altogether (Groot et al., 2021; Kucyi et al., 2017). To complement this approach, and in accordance with the results from the present study, the FT-RSGT is a good step in the right direction. The FT-RSGT was able to capture well-known task maskers such as time on task effects, as well as robust behavioral manifestations of MW with high temporal resolution.

Another problem with the current self-report measurements in the current version of the thought probes. The current use of the thought probes assumes that the content of thoughts is available for introspection. The current model uses task data from ~18 seconds before the thought probe appeared, and it is not unlikely that some participants might report an average of their thought focus for the past minute, therefore inaccurately matching their thought content to the behavioral responses in the time window in question (even though the question clearly states right before this question appeared). Additionally, the current version of the thought probes does not deal with different types of mind-wandering, if the participant does not have access to the content of their thoughts, or if they fail to classify their thoughts in the correct category according to the instructions given before the task.

Task comprehension

There is also the problem of equal task comprehension amongst participants. Even though task instructions are standardized, the concept of randomness might be difficult for some participants to understand, or interpreted different, especially when it comes to creating a random pattern with only two buttons. Some participants might be on task yet perform poorly according to the AE behavioral measurement simply because they used the wrong strategy. One way to solve the problem of task comprehension could be to incorporate feedback at some point during training or the task. In the current version of the FT-RSGT participants are only instructed how to solve the task but are not allowed insight into the algorithm calculating their performance. Incorporating a feedback system during training of the FT-RSGT could be promising approach when it comes to standardizing task comprehension.

An additional concern is the robustness of the effect when replicated in different cultures, languages, and labs. As discussed in the introduction, the original study by Axelrod et al. (2015) was done in Israel and their follow up replication in China (Axelrod et al., 2018). While they claimed to successfully replicate their original effect, the studies were criticized for being underpowered, and for using different stimulation montages thereby compromising the similarities between the original study and the replication (Csifcsák et al., 2018). The large-scale replication of Axelrod et al. (2015) by Boayue et al. (2019) was done in laboratories across 3 countries (Sweden, Norway, and the Netherlands), and 4 languages (Swedish, Norwegian, Dutch and English), which undoubtedly increases between subject variability. Clearly the true effect of tDCS on MW (should it exist) is not very large and should therefore be harder to detect when adding extra between subject variability such as language/cultural interpretations and differences. This is not a brain stimulation specific problem as researchers have noted that some problems can arise when attempting to generalize findings across cultures (e.g., Berry, 1969; Poortinga, 1989; Van De Vijver &

Poortinga, 1982). However, in terms of efficacy of tDCS on MW, the argument can be made that for tDCS to have any real-world applicability (such as treating mood disorders) the effect needs to be robust enough to withstand such differences.

Motivation

Motivation is an important factor in task engagement and sustained attention (Appel & Gilabert, 2002; Arnold, 1985; Fortenbaugh et al., 2017), and reward modulation can induce neuronal changes in sustained attention tasks (Esterman et al., 2017). MW research often uses monotonous tasks (the FTRSGT is no exception) to induce MW to avoid a ceiling effect in on-task performance. To investigate MW, the task needs to promote and give opportunity for the participants to MW, however one could argue that lack of motivation could produce a floor effect where participants just give up early in the task. The current study was set up without an extra motivation for performing well and relying on intrinsic problem-solving motivation which has been shown to differ greatly in individuals (Arnold, 1985). A great example of this is one participant in the current study used only one key during the entire task, and only responded “on task” to all 30 thought probes. Though we could not exclude this participant as per our exclusion criteria, this is a typical example of low motivation task performance. Losing valuable data sets in such a way can be problematic in a field which already has problems with expensive data collection and small samples.

A solution to this could be to pay participants a standard fee for participating and offer some extra money should they perform well (while still paying everyone the same amount of money), which has been used in research where motivation is important to manipulate (e.g., decision making research; Csifcsak, et al., 2019; Csifcsák et al., 2020). Motivation is a potential problem that should be accounted for (either checking for floor effects in the data, or by post experiment questionnaires). Apart from the one data set mentioned above, it seems

unlikely that it happened in this study, as we did observe an effect of decreased task performance both within and between parts, suggesting slowly declining task performance which is to be expected this type of research.

As we can see, there are a lot of concerns to address when designing and conducting brain stimulation research to investigate MW. It is up to the individual researcher and reviewer to make sure the quality of the research we produce is high enough to warrant spending the substantial amount of money that is poured into research every year, which leads us into other non-methodological problems with replicating research, funding and publication.

Non-methodological reasons for low replicability

Funding and publication bias

A large amount of money is annually poured into research, as an example, it is estimated that in biomedical research in the US alone consumes upwards of 250 billion dollars a year (Moher et al., 2016). The problem is that it is estimated that large amounts of that funding was being avoidably wasted (Chalmers & Glasziou, 2009). The National Institute of Health research in the United Kingdom proposed a five-stage model to reduce research waste. The model included steps such as increasing statistical power, replication of initial observations, bias reduction, under-reporting of disappointing results and biased data reporting within studies to mention a few (Chalmers & Glasziou, 2009; Moher et al., 2016). Running poor methodological studies that end up never getting published is both a waste of time and resources. Alternatively, publishing weak methodological studies is arguably worse, as false positives can inspire investment into more research to waste resources on (Simmons et al., 2011).

In addition to researchers inherently wanting to chase the new and innovative, there is also massive pressure to secure funding and to publish research as tenure and promotions are

determined by publication rates (Boice & Jones, 1984; Parchomovsky, 2000). Unfortunately, the chances of securing funding for new and innovative studies are higher than securing funding for replication research (Spector et al., 2015). Although steps are being taken to combat the inequality of funding (e.g., Baker, 2016), the inequality in funding can bias the academic literature towards novel studies. This inequality could also work as extra pressure on researchers to use researcher degrees of freedom in search of novel and exiting findings, which might be hard to replicate.

A lot of private funding also helps support research (Muscio et al., 2013), and to avoid any undisclosed bias it is common to report any conflict of interest in an own section when publishing a report. While private funding is critical for a lot of research, in some fields, private funding might also come with a "cost", such as funders reserving of the right to withhold results or specific information should they be deemed not to be "industry healthy" (Lexchin et al., 2003). A typical example of this could be a private pharmaceutical company that funds research where the results suggests that one of their products has no effect or has undesirable effects. In this case, the pharmaceutical company might want to discontinue the funding, or withhold the results from being published (Cohen, 2014; Lexchin et al., 2003; Lexchin, 2005). A review of pharmaceutical research found that industry funded research were twice as likely to not share research data or results, compared to those without industry backing (Lexchin, 2005). Additionally, commercially funded clinical research was more likely to yield positive results (especially pro-industry significant results), then compared with other sources of funding (Bhandari et al., 2004; Lexchin, 2005). Luckily, this is, to my knowledge, not common in psychological research, yet it might be too naive to believe commercial funding and industry backing is not somewhat transferable cross-disciplines. Therefore it is likely that some of the publication bias in the current psychological literature is partly due to conflict of interest in industry funded researchers (Harvey, 2011).

To combat publication bias, there is promise in the form of registered reports, and registered replications. Unfortunately, registering replications have "bad" reputation of "killing effects", which stems from the fact that a lot of registered replications fail to replicate the original effect (Open Science Collaboration, 2015). In brain stimulation research both Boayue et al. (2019) and the present study are good examples. Some researchers even suggest that when researchers get heavily invested in a field, disproving a beloved theory is less appealing than building upon it (Ferguson & Heene, 2012). This is even more true where there is more ambiguity in scientific theory (such as in psychology and psychiatry; Fanelli, 2010). As an example, over 90% of published results in psychology are theory supporting, while in other fields such as astronomy and physics only 70% of published reports are theory supporting (Fanelli, 2010). Researchers have pointed out that this discrepancy likely stems from the fact that social sciences are more flexible compared to some of the "hard" sciences such as space sciences (Ferguson & Heene, 2012).

Another step taken towards combating publication bias and under-reporting of dissatisfactory results, is an increased use of preprints and database initiatives such as the AllTrials Campaign (https://www.alltrials.net/). Some journals are also behind this movement by making replication studies free of charge to publish and setting higher demands for statistical power requirements to publish, as well as opening for accepting registered reports. Registered reports are a good way to get feedback on a study (the present study had two revisions of the stage-1 report), as well as guaranteeing that the final report will be published. Registering reports is also a good way to counter the file drawer problem (Rosenthal, 1979) as well as publication bias (Chambers, 2013; Lakens, 2019; Waters et al., 2020).

In addition to registering reports, the use of open access repositories is also a promising development when it comes to combating publication bias. Open access repositories make it easy to share raw research data, methods, and materials. This saves a lot of resources

(especially research where data collection is expensive), as well as facilitating replication and promotes transparent research. By sharing raw data in an open access repository, researchers can save both time and money by using the same dataset for multiple purposes (such as the example of Fassi & Kadosh, 2020). One example of this is using these freely available datasets instead of pilots to run power calculations before starting a new study. Sharing raw data in repositories as well as using preprint servers also ensure that not only published papers and final samples get published. Additionally using preprint servers increases the pool of available knowledge, and at the very least make failed studies accessible to prevent further resource waste. Lastly, and perhaps most importantly it also ensures that the literature does not only consist of significant findings that make it through the journal selection process.

In fact, there are many benefits in sharing research data (Borgman, 2012; Figueiredo et al., 2019). As mentioned by sharing analysis methods and raw data, the cost and time it takes to replicate an existing study is reduced, as well as making it easier for the study to be included in a meta-analysis (which in turn can increase the citation of the article; Piwowar et al., 2007). Registering studies and using open access repositories seems like the obvious choice when designing a study, however, there are also a few problems with sharing data, registering studies, and publishing all studies regardless of results. Perhaps the most obvious problem with registering is that it is time consuming. Registering a study, especially a registered report takes a lot of time. Often, studies are time limited due to academic time frames where there simply is not enough time or other resources to plan and implement a registered report. However, registering the final analysis pipeline does not take long, and should be a requirement of any confirmatory study that aims at getting published. Some researchers might also be afraid of their ideas being stolen by other researchers. As we examined earlier, researchers are under constant pressure to publish to advance their career, which creates a competitive environment where openly sharing the entire research process might not seem

tempting to everyone. Another issue with sharing research data is sensitive data. Not all raw data can be shared, and sensitive data will have to be anonymized before publicly shared, which again takes time and effort. However, this argument pales in comparison to the amount of time and resources that it can save.

Concluding remarks

This thesis presents evidence that HD-tDCS over the left DLPFC cannot reduce the propensity to MW, that the FT-RSGT can capture promising behavioral indices of MW, and it has highlighted the importance of thoroughly investigating novel findings with high powered, pre-registered study designs. Additionally, the thesis has reviewed some of the main methodological problems with brain stimulation research and proposed possible reasons for why brain stimulation research on mind wandering often fails to replicate. Finally, the thesis has discussed what can be done to further progress in brain stimulation research, and to an extent, psychological research as a field. And while novel exciting research is the driving force of scientific progress, I wish to end this thesis in the words of the Open Science Collaboration: “Innovation points out paths that are possible; replication points out paths that are likely; progress relies on both” (Open Science Collaboration, 2015).

References

Alltrials campaign. (n.d). Retrieved from 15. May 2022. https://www.alltrials.net/

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, *16*(1), 17–42. https://doi.org/10.1007/s11065-006-9002-x

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., & Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *Journal of Neurophysiology*, *104*(1), 322–335. https://doi.org/10.1152/jn.00830.2009

Antrobus, J. S., Singer, J. L., Goldstein, S., & Fortgang, M. (1970). Section of Psychology: MINDWANDERING AND COGNITIVE STRUCTURE. *Transactions of the New York Academy of Sciences*, *32*(2), 242–252. https://doi.org/10.1111/j.2164-0947.1970.tb02056.x

Appel, C., & Gilabert, R. (2002). Motivation and task performance in a task-based web-based tandem project. *ReCALL*, *14*(1), 16–31. https://doi.org/10.1017/S0958344002000319

Arnold, H. J. (1985). Task performance, perceived competence, and attributed causes of performance as determinants of intrinsic motivation. *Academy of Management Journal*, *28*(4), 876–888. https://doi.org/10.5465/256242

Asendorpf, J. B., Conner, M., de Fruyt, F., de Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Replication is more than hitting the lottery twice. *European Journal of Personality*, *27*(2), 138–144.

Asendorpf, J. B., Conner, M., de Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2016).

Recommendations for increasing replicability in psychology (p. 622). American Psychological Association. https://doi.org/10.1037/14805-038

Axelrod, V., Rees, G., Lavidor, M., & Bar, M. (2015). Increasing propensity to mind-wander with transcranial direct current stimulation. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3314–3319. PsycINFO. https://doi.org/10.1073/pnas.1421435112

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. https://doi.org/10.1038/s41598-018-34098-z

Baddeley, A. D., & Hitch, G. (1974). Working Memory. I G. H. Bower (Red.), *Psychology of Learning and Motivation* (Bd. 8, s. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

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. https://doi.org/10.1080/713755788

Baker, M. (2016). Dutch agency launches first grants programme dedicated to replication. *Nature*. https://doi.org/10.1038/nature.2016.20287

Begg, C. B., & Berlin, J. A. (1989). Publication bias and dissemination of clinical research. *JNCI: Journal of the National Cancer Institute*, *81*(2), 107–115. https://doi.org/10.1093/jnci/81.2.107

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, *4*(2), 119–128. https://doi.org/10.1080/00207596908247261

Bhandari, M., Busse, J. W., Jackowski, D., Montori, V. M., Schünemann, H., Sprague, S., Mears, D., Schemitsch, E. H., Heels-Ansdell, D., & Devereaux, P. J. (2004).

Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *CMAJ*, *170*(4), 477–480.

Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*, *90*(2), 196–208. https://doi.org/10.1016/j.physbeh.2006.09.009

Boayue, N. M., 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. *European Journal of Neuroscience*, *0*(0). https://doi.org/10.1111/ejn.14347

Boayue, N. M., Csifcsák, G., Kreis, I. V., Schmidt, C., Finn, I., Hovde Vollsund, A. E., & Mittner, M. (2020). The interplay between executive control, behavioural variability and mind wandering: Insights from a high-definition transcranial direct-current stimulation study. *European Journal of Neuroscience*, *n/a*(n/a). https://doi.org/10.1111/ejn.15049

Boice, R., & Jones, F. (1984). Why academicians don't write. *Journal of Higher Education*, *55*(5), 567–582. https://doi.org/10.2307/1981822

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*(3), 333–342. https://doi.org/10.1177/1745691614529796

Brick, J., & Erickson, C. K. (2012). *Drugs, the brain, and behavior: The pharmacology of drug use disorders* (2. edition). Routledge. https://doi.org/10.4324/9780203075814

Broadway, J. M., Zedelius, C. M., Mooneyham, B. W., Mrazek, M. D., & Schooler, J. W. (2015). Stimulating minds to wander. *Proceedings of the National Academy of Sciences*, *112*(11), 3182–3183. https://doi.org/10.1073/pnas.1503093112

Brunoni, A. R., & Fregni, F. (2011). Clinical trial design in non-invasive brain stimulation psychiatric research. *International Journal of Methods in Psychiatric Research*, *20*(2), e19–e30. https://doi.org/10.1002/mpr.338

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA*, *287*(21), 2847–2850. https://doi.org/10.1001/jama.287.21.2847

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98–113. https://doi.org/10.1016/j.cognition.2008.12.009

Chiesa, A., & Serretti, A. (2009). Mindfulness-based stress reduction for stress management in healthy people: A Review and Meta-Analysis. *The Journal of Alternative and Complementary Medicine*, *15*(5), 593–600. https://doi.org/10.1089/acm.2008.0495

Chou, T., Hooley, J. M., & Camprodon, J. A. (2020). Transcranial direct current stimulation of default mode network parietal nodes decreases negative mind-wandering about the past. *Cognitive Therapy and Research*, *44*(1), 10–20. https://doi.org/10.1007/s10608-019-10044-9

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, *17*(11), 718–731. https://doi.org/10.1038/nrn.2016.113

Christoff, K., Mills, C., Andrews-Hanna, J. R., Irving, Z. C., Thompson, E., Fox, K. C. R., & Kam, J. W. Y. (2018). Mind-wandering as a scientific concept: Cutting through the definitional haze. *Trends in Cognitive Sciences*, *22*(11), 957–959. https://doi.org/10.1016/j.tics.2018.07.004

Cohen, D. (2014). Dabigatran: How the drug company withheld important analyses. *BMJ*, *349*, g4670. https://doi.org/10.1136/bmj.g4670

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Cole, J. C., Green Bernacki, C., Helmer, A., Pinninti, N., & O'reardon, J. P. (2015). Efficacy of transcranial magnetic stimulation (TMS) in the treatment of Schizophrenia: A review of the literature to date. *Innovations in Clinical Neuroscience*, *12*(7–8), 12–19.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Coulborn, S., Bowman, H., Miall, R. C., & Fernández-Espejo, D. (2020). Effect of tDCS over the right inferior parietal lobule on mind-wandering propensity. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00230

Csifcsak, G., Boayue, N. M., Aslaksen, P. M., Turi, Z., Antal, A., Groot, J., Hawkins, G. E., Forstmann, B. U., Opitz, A., Thielscher, A., & Mittner, M. (2019). Commentary: Transcranial stimulation of the frontal lobes increases propensity of mind-wandering without changing meta-awareness. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00130

Csifcsák, G., Boayue, N. M., Puonti, O., Thielscher, A., & Mittner, M. (2018). Effects of transcranial direct current stimulation for treating depression: A modeling study. *Journal of Affective Disorders*, *234*, 164–173. https://doi.org/10.1016/j.jad.2018.02.077

Csifcsak, G., Melsaeter, E., & Mittner, M. (2019). Intermittent absence of control during reinforcement learning interferes with Pavlovian bias in action selection. *J Cogn Neurosci*, 1–18. https://doi.org/10.1162/jocn_a_01515

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*(7), 376–385. https://doi.org/10.1177/070674371305800702

Curtiss, J. H. (1940). Review of mathematics of statistics. *The American Mathematical Monthly*, *47*(5), 309–311. https://doi.org/10.2307/2302695

Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., & Bikson, M. (2009). Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimulation*, *2*(4), 201-207.e1. https://doi.org/10.1016/j.brs.2009.03.005

Datta, A., Elwassif, M., Battaglia, F., & Bikson, M. (2008). Transcranial current stimulation focality using disc and ring electrode configurations: FEM analysis. *Journal of Neural Engineering*, *5*(2), 163–174. https://doi.org/10.1088/1741-2560/5/2/007

deCharms, R. C. (2007). Reading and controlling human brain activation using real-time functional magnetic resonance imaging. *Trends in Cognitive Sciences*, *11*(11), 473–481. https://doi.org/10.1016/j.tics.2007.08.014

Dixon, M., Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, *115*, 201715766. https://doi.org/10.1073/pnas.1715766115

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53–78. https://doi.org/10.1348/000712601162103

Drotar, D. (2010). Editorial: A call for replications of research in pediatric psychology and guidance for authors. *Journal of Pediatric Psychology*, *35*(8), 801–805. https://doi.org/10.1093/jpepsy/jsq049

Esterman, M., Poole, V., Liu, G., & DeGutis, J. (2017). Modulating reward induces differential neurocognitive approaches to sustained attention. *Cerebral Cortex*, *27*(8), 4022–4032. https://doi.org/10.1093/cercor/bhw214

Evans, C., Bachmann, C., Lee, J. S. A., Gregoriou, E., Ward, N., & Bestmann, S. (2020). Dose-controlled tDCS reduces electric field intensity variability at a cortical target site. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *13*(1), 125–136. https://doi.org/10.1016/j.brs.2019.10.004

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLOS ONE*, *5*(4), e10271. https://doi.org/10.1371/journal.pone.0010271

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Fassi, L., & Kadosh, R. C. (2020). Is it all in our head? When subjective beliefs about receiving an intervention are better predictors of experimental results than the intervention itself (s. 2020.12.06.411850). bioRxiv. https://doi.org/10.1101/2020.12.06.411850

Ferguson, C. J. (2009). Is psychological research really as good as medical research? Effect size comparisons between psychology and medicine. *Review of General Psychology, 13*(2), 130–136. https://doi.org/10.1037/a0015103

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science, 7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Figueiredo Filho, D., Lins, R., Domingos, A., Janz, N., Silva, L., Figueiredo Filho, D., Lins, R., Domingos, A., Janz, N., & Silva, L. (2019). Seven reasons why: A user's guide to transparency and reproducibility. *Brazilian Political Science Review*, *13*(2). https://doi.org/10.1590/1981-3821201900020001

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. https://doi.org/10.1016/j.neuropsychologia.2019.04.013

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. https://doi.org/10.1016/j.neuropsychologia.2020.107723

Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, *1396*(1), 70–91. https://doi.org/10.1111/nyas.13318

Fox, K. C. R., Nijeboer, S., Solomonova, E., Domhoff, G. W., & Christoff, K. (2013). Dreaming as mind wandering: Evidence from functional neuroimaging and first-person content reports. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00412

Fox, K. C. R., 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. https://doi.org/10.1016/j.neuroimage.2015.02.039

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), 72–89. https://doi.org/10.1038/s41386-021-01132-0

Gordon, M., Dux, P., & Filmer, H. (2021). Not quite all in our head: Intervention is a better predictor of tDCS mind-wandering effects than subjective beliefs about experimental result. bioRxiv. https://doi.org/10.1101/2021.11.14.468553

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*(1), 1–20. https://doi.org/10.1037/h0076157

Groot, J. M., Boayue, N. M., Csifcsák, G., Boekel, W., Huster, R., Forstmann, B. U., & Mittner, M. (2021). Probing the neural signature of mind wandering with simultaneous fMRI-EEG and pupillometry. *NeuroImage*, *224*, 117412. https://doi.org/10.1016/j.neuroimage.2020.117412

Groot, J. M., Csifcsák, G., Wientjes, S., Forstmann, B. U., & Mittner, M. (2022). Catching wandering minds with tapping fingers: Neural and behavioral insights into task-unrelated cognition. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhab494

Gross, R. G., & Grossman, M. (2010). EXECUTIVE RESOURCES. *Continuum (Minneapolis, Minn.)*, *16*(4 0 0), 140–152. https://doi.org/10.1212/01.CON.0000368266.46038.0e

Hamra, G. B., MacLehose, R. F., & Cole, S. R. (2013). Sensitivity analyses for sparse-data problems—using weakly informative bayesian priors. *Epidemiology (Cambridge, Mass.)*, *24*(2), 233–239. https://doi.org/10.1097/EDE.0b013e318280db1d

Han, D. H., Hwang, J. W., & Renshaw, P. F. (20110808). Bupropion sustained release treatment decreases craving for video games and cue-induced brain activity in patients with internet video game addiction. *Psychology of Popular Media Culture*, *1*(S), 108. https://doi.org/10.1037/2160-4134.1.S.108

Harvey, C. R., & Liu, Y. (2021). Uncovering the iceberg from its tip: A model of publication bias and p-hacking (SSRN Scholarly Paper Nr. 3865813). Social Science Research Network. https://doi.org/10.2139/ssrn.3865813

Harvey, E. J. (2011). Research funded by the industry. *Canadian Journal of Surgery*, *54*(5), 293. https://doi.org/10.1503/cjs.023811

Hordacre, B. (2018). The role of telehealth to assist in-home tDCS: Opportunities, promising results and acceptability. *Brain Sciences*, *8*(6), 102. https://doi.org/10.3390/brainsci8060102

Horvath, J. C., Forte, J. D., & Carter, O. (2015). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP amplitude modulation in healthy human subjects: A systematic review. *Neuropsychologia*, *66*, 213–236. https://doi.org/10.1016/j.neuropsychologia.2014.11.021

Huang, Y., Liu, A. A., Lafon, B., Friedman, D., Dayan, M., Wang, X., Bikson, M., Doyle, W. K., Devinsky, O., & Parra, L. C. (2017). Measurements and models of electric fields in

the in vivo human brain during transcranial electric stimulation. *eLife*, *6*, e18834. https://doi.org/10.7554/eLife.18834

JASP Team. (2021). *(Version 0.9) [Computer software]*. https://jasp-stats.org/

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, *23*, 524–532. https://doi.org/10.1177/0956797611430953

Johnson, K. A., Robertson, I. H., Kelly, S. P., Silk, T. J., Barry, E., Dáibhis, A., Watchorn, A., Keavey, M., Fitzgerald, M., Gallagher, L., Gill, M., & Bellgrove, M. A. (2007). Dissociation in performance of children with ADHD and high-functioning autism on a task of sustained attention. *Neuropsychologia*, *45*(10), 2234–2245. https://doi.org/10.1016/j.neuropsychologia.2007.02.019

Jubera-Garcia, E., Gevers, W., & Van Opstal, F. (2021). Local build-up of sleep pressure could trigger mind wandering: Evidence from sleep, circadian and mind wandering research. *Biochemical Pharmacology*, *191*, 114478. https://doi.org/10.1016/j.bcp.2021.114478

Kajimura, S., & Nomura, M. (2015). Decreasing propensity to mind-wander with transcranial direct current stimulation. *Neuropsychologia*, *75*, 533–537. https://doi.org/10.1016/j.neuropsychologia.2015.07.013

Kam, J. W. Y., & Handy, T. C. (2014). Differential recruitment of executive resources during mind wandering. *Consciousness and Cognition*, *26*, 51–63. https://doi.org/10.1016/j.concog.2014.03.002

Kam, J. W. Y., Mittner, M., & Knight, R. T. (2022). Mind-wandering: Mechanistic insights from lesion, tDCS, and iEEG. *Trends in Cognitive Sciences*, *26*(3), 268–282. https://doi.org/10.1016/j.tics.2021.12.005

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. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439

Klooster, D. C. W., de Louw, A. J. A., Aldenkamp, A. P., Besseling, R. M. H., Mestrom, R. M. C., Carrette, S., Zinger, S., Bergmans, J. W. M., Mess, W. H., Vonck, K., Carrette, E., Breuer, L. E. M., Bernas, A., Tijhuis, A. G., & Boon, P. (2016). Technical aspects of neurostimulation: Focus on equipment, electric field modeling, and stimulation protocols. *Neuroscience & Biobehavioral Reviews*, *65*, 113–141. https://doi.org/10.1016/j.neubiorev.2016.02.016

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral Cortex*, *27*(3), 1831–1840. https://doi.org/10.1093/cercor/bhw029

Kuo, M.-F., & Nitsche, M. A. (2012). Effects of transcranial electrical stimulation on cognition. *Clinical EEG and Neuroscience*, *43*(3), 192–199. https://doi.org/10.1177/1550059412444975

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. https://doi.org/10.31234/osf.io/jbh4w

Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., Cantello, R. M., Cincotta, M., de Carvalho, M., De Ridder, D., Devanne, H., Di Lazzaro, V., Filipović, S. R., Hummel, F. C., Jääskeläinen, S. K., Kimiskidis, V. K., Koch, G., Langguth, B., Nyffeler, T., … Garcia-Larrea, L. (2014). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS). *Clinical Neurophysiology*, *125*(11), 2150–2206. https://doi.org/10.1016/j.clinph.2014.05.021

Levkovitz, Y., Isserles, M., Padberg, F., Lisanby, S. H., Bystritsky, A., Xia, G., Tendler, A., Daskalakis, Z. J., Winston, J. L., Dannon, P., Hafez, H. M., Reti, I. M., Morales, O. G., Schlaepfer, T. E., Hollander, E., Berman, J. A., Husain, M. M., Sofer, U., Stein, A., … Zangen, A. (2015). Efficacy and safety of deep transcranial magnetic stimulation for major depression: A prospective multicenter randomized controlled trial. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *14*(1), 64–73. https://doi.org/10.1002/wps.20199

Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *BMJ*, *326*(7400), 1167–1170. https://doi.org/10.1136/bmj.326.7400.1167

Lexchin, J. R. (2005). Implications of pharmaceutical industry funding on clinical research. *Annals of Pharmacotherapy*, *39*(1), 194–197. https://doi.org/10.1345/aph.1E224

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. https://doi.org/10.1016/j.jesp.2018.08.009

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*(12), 1827–1832. https://doi.org/10.1177/0956797615616374

Loo, C. K., & Mitchell, P. B. (2005). A review of the efficacy of transcranial magnetic stimulation (TMS) treatment for depression, and current and future strategies to optimize efficacy. *Journal of Affective Disorders*, *88*(3), 255–267. https://doi.org/10.1016/j.jad.2005.08.001

Luo, Y., Weibman, D., Halperin, J. M., & Li, X. (2019). A review of heterogeneity in Attention Deficit/Hyperactivity Disorder (ADHD). *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/article/10.3389/fnhum.2019.00042

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*(3), 679–690. https://doi.org/10.3758/s13428-010-0049-5

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. https://doi.org/10.1080/10705511.2016.1186549

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? *Psychological bulletin*, *136*(2), 188–207. https://doi.org/10.1037/a0018298

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin & Review*, *16*(5), 857–863. https://doi.org/10.3758/PBR.16.5.857

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moher, D., Glasziou, P., Chalmers, I., Nasser, M., Bossuyt, P. M. M., Korevaar, D. A., Graham, I. D., Ravaud, P., & Boutron, I. (2016). Increasing value and reducing waste in biomedical research: Who's listening? *The Lancet*, *387*(10027), 1573–1586. https://doi.org/10.1016/S0140-6736(15)00307-4

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs*. http://CRAN.R-project.org/

Muscio, A., Quaglione, D., & Vallanti, G. (2013). Does government funding complement or substitute private research funding to universities? *Research Policy*, *42*(1), 63–75. https://doi.org/10.1016/j.respol.2012.04.010

Nathan, S. S., Sinha, S. R., Gordon, B., Lesser, R. P., & Thakor, N. V. (1993). Determination of current density distributions generated by electrical stimulation of the human cerebral cortex. *Electroencephalography and Clinical Neurophysiology*, *86*(3), 183–192. https://doi.org/10.1016/0013-4694(93)90006-H

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., Paulus, W., Hummel, F., Boggio, P. S., Fregni, F., & Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, *1*(3), 206–223. https://doi.org/10.1016/j.brs.2008.06.004

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, *57*(10), 1899–1901. https://doi.org/10.1212/wnl.57.10.1899

Ocasio, W. (2011). Attention to attention. *Organization Science*, *22*(5), 1286–1296.

Ottaviani, C., & Couyoumdjian, A. (2013). Pros and cons of a wandering mind: A prospective study. *Front Psychol*, *4*, 524. https://doi.org/10.3389/fpsyg.2013.00524

Parchomovsky, G. (2000). Publish or perish. *Michigan Law Review*, *98*(4), 926–952. https://doi.org/10.2307/1290335

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612465253

Peñalva, R. G., Lancel, M., Flachskamm, C., Reul, J. M. H. M., Holsboer, F., & Linthorst, A. C. E. (2003). Effect of sleep and sleep deprivation on serotonergic neurotransmission in the hippocampus: A combined in vivo microdialysis/EEG study in rats. *European Journal of Neuroscience*, *17*(9), 1896–1906. https://doi.org/10.1046/j.1460-9568.2003.02612.x

Pincus, S., & Kalman, R. E. (1997). Not all (possibly) "random" sequences are created equal. *Proceedings of the National Academy of Sciences*, *94*(8), 3513–3518. https://doi.org/10.1073/pnas.94.8.3513

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, *88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Pincus, S., & Singer, B. H. (1996). Randomness and degrees of irregularity. *Proceedings of the National Academy of Sciences*, *93*(5), 2083–2088. https://doi.org/10.1073/pnas.93.5.2083

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, *2*(3), e308. https://doi.org/10.1371/journal.pone.0000308

Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology*, *24*(6), 737–756. https://doi.org/10.1080/00207598908247842

Pribram, K. H. (1973). The primate frontal cortex: Executive of the brain. I *Psychophysiology of the frontal lobes* (s. xii, 332–xii, 332). Academic Press.

Purpura, D. P., & McMurtry, J. G. (1965). Intracellular activities and evoked potential changes during polarization of motor cortex. *Journal of Neurophysiology*, *28*(1), 166–185. https://doi.org/10.1152/jn.1965.28.1.166

R Core Team. (2015). *R: A language and environment for statistical com-puting.* https://www.R-project.org/.

Rahman, A., Reato, D., Arlotti, M., Gasca, F., Datta, A., Parra, L. C., & Bikson, M. (2013). Cellular effects of acute direct current stimulation: Somatic and synaptic terminal effects. *The Journal of Physiology*, *591*(10), 2563–2578. https://doi.org/10.1113/jphysiol.2012.247171

Rasmussen, I. D., Boayue, N. M., Mittner, M., Bystad, M., Grønli, O. K., Vangberg, T. R., Csifcsák, G., & Aslaksen, P. M. (2021). High-definition transcranial direct current stimulation improves delayed memory in Alzheimer's disease patients: A pilot study using computational modeling to optimize electrode position. *Journal of Alzheimer's Disease*, *83*(2), 753–769. https://doi.org/10.3233/JAD-210378

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rubio, B., Boes, A. D., Laganiere, S., Rotenberg, A., Jeurissen, D., & Pascual-Leone, A. (2016). Noninvasive brain stimulation in pediatric attention-deficit hyperactivity disorder (ADHD): A Review. *Journal of Child Neurology*, *31*(6), 784–796. https://doi.org/10.1177/0883073815615672

Sanborn, A. N., Hills, T. T., Dougherty, M. R., Thomas, R. P., Yu, E. C., & Sprenger, A. M. (2014). Reply to Rouder (2014): Good frequentist properties raise confidence. *Psychonomic Bulletin & Review*, *21*(2), 309–311. https://doi.org/10.3758/s13423-014-0607-4

Savin, N. E. (1980). The Bonferroni and the Scheffé multiple comparison procedures. *The Review of Economic Studies*, *47*(1), 255–273. https://doi.org/10.2307/2297112

Seibt, O., Brunoni, A. R., Huang, Y., & Bikson, M. (2015). The pursuit of DLPFC: Non-neuronavigated methods to target the left dorsolateral pre-frontal cortex with symmetric bicephalic transcranial direct current stimulation (tDCS). *Brain Stimulation*, *8*(3), 590–602. https://doi.org/10.1016/j.brs.2015.01.401

Seli, P., Carriere, J., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00430

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. https://doi.org/10.1037/a0030954

Seli, P., Jonker, T., Cheyne, J., & Smilek, D. (2013). Enhancing SART validity by statistically controlling speed-accuracy trade-offs. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00265

Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences*, *22*(6), 479–490. https://doi.org/10.1016/j.tics.2018.03.010

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simon, L., & Bikson, M. (2019). The frontier of tDCS in psychiatry and the role of new technologies. *L'Encephale*, *45 Suppl 2*, S55–S57. https://doi.org/10.1016/j.encep.2019.04.004

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2019). Specification curve: Descriptive and inferential statistics on all reasonable specifications (SSRN Scholarly Paper Nr. 2694998). Social Science Research Network. https://doi.org/10.2139/ssrn.2694998

Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *The Cognitive Neuroscience of Thought*, *1428*, 60–70. https://doi.org/10.1016/j.brainres.2011.03.072

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

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Spector, J. M., Johnson, T. E., & Young, P. A. (2015). An editorial on replication studies and scaling up efforts. *Educational Technology Research and Development*, *63*(1), 1–4. https://doi.org/10.1007/s11423-014-9364-3

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*(1), 303–317. https://doi.org/10.1016/j.neuroimage.2010.06.016

Stan Development Team. (2016). *RStan: The R interface to Stan*. http:// mc-stan.org/

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, *57*(3), 748–761. https://doi.org/10.1111/ajps.12001

Tarver, J., Daley, D., & Sayal, K. (2014). Attention-deficit hyperactivity disorder (ADHD): An updated review of the essential facts. *Child: Care, Health and Development*, *40*(6), 762–774. https://doi.org/10.1111/cch.12139

Teasdale, J. D., Segal, Z., & Williams, J. M. G. (1995). How does cognitive therapy prevent depressive relapse and why should attentional control (mindfulness) training help? *Behaviour Research and Therapy*, *33*(1), 25–39. https://doi.org/10.1016/0005-7967(94)E0011-7

Tortella, G., M.L. Selingardi, P., L. Moreno, M., P. Veronezi, B., & R. Brunoni, A. (2014). Does non-invasive brain stimulation improve cognition in major depressive disorder? A systematic review. *CNS & Neurological Disorders - Drug Targets- CNS & Neurological Disorders)*, *13*(10), 1759–1769.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology*, *89*(1), 77–101. https://doi.org/10.1111/j.2044-8295.1998.tb02674.x

Tremblay, S., Lepage, J.-F., Latulipe-Loiselle, A., Fregni, F., Pascual-Leone, A., & Théoret, H. (2014). The uncertain outcome of prefrontal tDCS. *Brain Stimulation*, *7*(6), 773–783. https://doi.org/10.1016/j.brs.2014.10.003

Turi, Z., Csifcsák, G., Boayue, N. M., Aslaksen, P., Antal, A., Paulus, W., Groot, J., Hawkins, G. E., Forstmann, B., Opitz, A., Thielscher, A., & Mittner, M. (2019). Blinding is compromised for transcranial direct current stimulation at 1 mA for 20 min in young healthy adults. *European Journal of Neuroscience*, *50*(8), 3261–3268. https://doi.org/10.1111/ejn.14403

Van De Vijver, F. J. R., & Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology*, *13*(4), 387–408. https://doi.org/10.1177/0022002182013004001

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology. General*, *143*(4), 1457–1475. https://doi.org/10.1037/a0036731

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Waters, A. M., LeBeau, R. T., Young, K. S., Dowell, T. L., & Ryan, K. M. (2020). Towards the enhancement of quality publication practices in clinical psychological science. *Behaviour Research and Therapy*, *124*, 103499. https://doi.org/10.1016/j.brat.2019.103499

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*(2), 163–206. https://doi.org/10.1037/0033-2909.134.2.163

Westwood, S. J., Radua, J., & Rubia, K. (2021). Noninvasive brain stimulation in children and adults with attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Journal of Psychiatry and Neuroscience*, *46*(1), E14–E33. https://doi.org/10.1503/jpn.190179

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01832

Yanko, M. R., & Spalek, T. M. (2013). Route familiarity breeds inattention: A driving simulator study. *Accident Analysis & Prevention*, *57*, 80–86. https://doi.org/10.1016/j.aap.2013.04.003

Zrenner, C., Desideri, D., Belardinelli, P., & Ziemann, U. (2018). Real-time EEG-defined excitability states determine efficacy of TMS-induced plasticity in human motor cortex. *Brain Stimulation*, *11*(2), 374–389. https://doi.org/10.1016/j.brs.2017.11.016

Åkerstedt, T. (2007). Altered sleep/wake patterns and mental performance. *Physiology & Behavior*, *90*(2), 209–218. https://doi.org/10.1016/j.physbeh.2006.09.007