Cortex

CrossMark

Research report

No evidential value in samples of transcranial direct current stimulation (tDCS) studies of cognition and working memory in healthy populations

*Jared Medina** and Samuel Cason*

Department of Psychological and Brain Sciences, University of Delaware, USA

ABSTRACT

A substantial number of studies have been published over the last decade, claiming that transcranial direct current stimulation (tDCS) can influence performance on cognitive tasks. However, there is some skepticism regarding the efficacy of tDCS, and evidence from meta-analyses are mixed. One major weakness of these meta-analyses is that they only examine outcomes in published studies. Given biases towards publishing positive results in the scientific literature, there may be a substantial "file-drawer" of unpublished negative results in the tDCS literature. Furthermore, multiple researcher degrees of freedom can also inflate published p-values. Recently, Simonsohn, Nelson and Simmons (2014) created a novel meta-analytic tool that examines the distribution of significant p-values in a literature, and compares it to expected distributions with different effect sizes. Using this tool, one can assess whether the selected studies have evidential value. Therefore, we examined a random selection of studies that used tDCS to alter performance on cognitive tasks, and tDCS studies on working memory in a recently published meta-analysis (Mancuso et al., 2016). Using a p-curve analysis, we found no evidence that the tDCS studies had evidential value (33% power or greater), with the estimate of statistical power of these studies being approximately 14% for the cognitive studies, and 5% (what would be expected from randomly generated data) for the working memory studies. It is likely that previous tDCS studies are substantially underpowered, and we provide suggestions for future research to increase the evidential value of future tDCS studies.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Transcranial direct current stimulation (tDCS) is an afford-able, non-invasive technique used to electrically stimulate the brain. Typically, two electrode pads (a positively charged anode and a negatively charged cathode) are placed on the participant. A relatively weak current (typically 1—2 mA) then runs from the cathode to the anode. This current is thought to change the resting membrane potential of neurons, resulting

in hyperpolarization (less activity) under the cathode, and hypopolarization (more activity) under the anode (Bindman, Lippold, & Redfearn, 1962; Nitsche & Paulus, 2000), along with long-term potentiation/depression-like plasticity after (respectively) anodal/cathodal stimulation (Stagg & Nitsche, 2011). This technique has been used to study changes in motor cortex excitability and motor learning (Nitsche et al., 2003). In addition to research on motor processes, tDCS has more recently been applied to a number of other domains, including altering cognitive (Sparing, Dafotakis, Meister, Thirugnanasambandam, & Fink, 2008) and emotional (Boggio, Zaghi, & Fregni, 2009) function in typical populations, as a therapy for stroke sufferers (Fregni, Boggio, Mansur, et al., 2005) and individuals with mental illness (Boggio et al., 2008), and to augment athletic training in high-level athletes (Banissy & Muggleton, 2013). In addition, it is inexpensive (with commercially available devices costing less than $200) and safe (Bikson et al., 2016). If such a simple device can be used to improve performance in all of these domains, its application could revolutionize brain science and rehabilitation. Therefore, it is of critical importance that claims regarding its effectiveness be examined and scrutinized.

There has been a substantial increase in the number of published tDCS studies, including studies of tDCS and cognitive processes, over the last five years.[1] Intuitively, the sheer number of manuscripts claiming an effect of tDCS on cognitive processes would suggest that this method clearly modulates behavior. However, there are multiple reasons why the number of published papers in a field is not always indicative of evidential value.[2] First, null results are typically not submitted for publication (the "file-drawer problem", Rosenthal, 1979), as editors are more likely to accept positive versus null results (Franco, Malhotra, & Simonovits, 2014). Although there are several studies showing a significant effect of tDCS on cognitive processes, there could be a larger number of unpublished studies that found no effect.

Second, decisions made during data analysis can falsely inflate significance. For example, researchers often have a number of decisions to make when collecting and analyzing data, including deciding on how many participants to test, whether (and how) to remove outliers, data transformations (e.g., using raw *vs* percentage scores, whether to normalize, etc.), which dependent variables should be reported or analyzed, whether to include covariates, whether to use median splits, type of statistical analysis to use, etc. Although it is best practice to decide on the analysis pipeline before data collection, these decisions can be made during data analysis and lead to potential biases (Gelman & Loken, 2013, 2014; Kunda, 1990). Furthermore, researchers may generate hypotheses after, not before, testing (HARKing: Hypothesizing After the Results are Known − see Kerr, 1998). For example, researchers may initially hypothesize that a specific

manipulation influences task performance. Not finding the predicted effect, the researcher can probe the data to examine if dividing the population into subsets (e.g., sex differences, median splits on a different variable) results in a significant effect of the manipulation. HARKing, using multiple analysis pipelines, and other practices (such as adding participants until a significant outcome is reached) all significantly increase the odds of a false positive finding (see Simmons, Nelson, & Simonsohn, 2011).

Third, underpowered studies are at risk of, not only false negatives (Type II errors) but also false positives (Type I errors) and overestimation of true effect sizes (Type M errors, see Gelman & Carlin, 2014). Small, underpowered studies can only detect large effects. If the true effect size is small or nonexistent, only studies that overestimate the true effect size via randomness will cross significance thresholds (the "winner's curse", see Button et al., 2013; Ioannidis, 2008). Therefore, a number of factors can lead to a substantial literature with limited evidential value.

Given these concerns, an important question is how to assess evidential value in the literature. Simonsohn, Nelson, and Simmons (2014b) have developed a method for testing the evidential value of a literature by examining reported p-values. Using this method, one first finds the distribution of significant ($p < .05$) p-values in a selection of published studies, ignoring any p-values that are not statistically significant. Next, one compares this distribution of p-values from the selected literature to distributions that would be expected given different effect sizes. For example, the distribution of p-values from a series of studies with no effect is expected to be flat, such that the same number of p-values should be observed between .12 and .13 or .74 and .75. Importantly, this is also true for significant p-values. If there is no true effect ($d = 0$), then there should be the same number of p-values from .01 to .02 as there are from .04 to .05. In the presence of a real effect, this p-value distribution should be right skewed, such that there are more observed p-values between .00 and .01 than between .04 and .05. On the other hand, given certain questionable research practices, researchers may stop collecting data or do exploratory analyses once they have crossed the critical $p < .05$ boundary. This practice, at times called p-hacking, would result in a distribution of p-values with left skew (more p-values closer to .05 than .00).

Previous papers have examined the effects of tDCS on various aspects of cognition using traditional meta-analysis techniques, with varying results. In the working memory domain, Brunoni and Vanderhasselt (2014) found that tDCS led to improvements in reaction time, but not accuracy. Hill, Fitzgerald, and Hoy (2016) found a small effect of anodal tDCS on offline (but not online) reaction time, while Mancuso, Ilieva, Hamilton, & Farah, 2016 found that left dorsolateral prefrontal cortex (DLPFC) tDCS improved performance when paired with training, with no other meta-analyses being significant. Dedoncker, Brunoni, Baeken, and Vanderhasselt (2016) reported that anodal, but not cathodal, DLPFC stimulation altered performance on cognitive tasks, whereas Horvath, Forte and Carter (Horvath, Forte, & Carter, 2015) found no evidence that tDCS influenced performance on cognitive tasks (though see Price & Hamilton, 2015 for discussion). However, Price, McAdams, Grossman, and Hamilton (2015)

---

[1] A Pubmed search for the terms tDCS or "transcranial direct current stimulation" found 68 manuscripts/year (12.8 that included "cognitive" or "cognition") from 2006 to 2010, and 370 manuscripts/year (115.8 with "cognitive" or "cognition") from 2011 to 2015.

[2] Evidential value − that the reported findings are due to tDCS influencing a cognitive process, and not due to some other non-evidential factor.

selected the language studies from Horvath, Forte & Carter, and reported that tDCS can influence performance on language tasks – though this finding has been questioned as well (see Westwood, Olson, Miall, Nappo, & Romani, 2017).

One issue with traditional meta-analytic techniques is that they can only be done on available data. Given publication biases for reporting significant results throughout the scientific literature, meta-analyses that pull from the published literature will be made primarily (or solely) of significant results, and are not able to take into account unpublished null findings. The most commonly used method to correct for publication bias is Trim and Fill (Duval & Tweedie, 2000a, 2000b), in which one examines a funnel plot for asymmetry, "trims" the studies with the largest effect sizes until symmetry is obtained, and then replaces the trimmed studies maintaining a symmetric funnel plot. However, Trim and Fill has been found to be inadequate at detecting publication bias. Simonsohn, Nelson, and Simmons (2014a) simulated datasets with varying true effect sizes (from $d = .0$ to $.8$), and then examined the estimated effect sizes from Trim and Fill compared to the p-curve analysis. They found that Trim and Fill vastly overestimated the effect size, such that Trim and Fill meta-analyses reported an effect size of about $.7$ from datasets with a true effect size of zero. In contrast, p-curve analyses provided an accurate estimate of the true effect size.

Given its advantages, we used a p-curve analysis to examine the evidential value of tDCS studies of cognition. This method has been used both to demonstrate that certain sets of studies have evidential value (e.g., studies of syntactic priming, see Mahowald, James, Futrell, & Gibson, 2016) and others do not (e.g., studies on "far transfer" in working memory with active controls and "power posing", see Melby-Lervåg, Redick, & Hulme, 2016; Simmons & Simonsohn, 2017). Therefore, we conducted two sets of meta-analyses to examine the effect of tDCS on cognitive processes. In our first set, we randomly selected 30 studies that could be included in the p-curve analysis that examined the effect of tDCS on cognition. Then, to account for potential criticisms of the first meta-analysis (study heterogeneity, possibilities of a biased sample), we conducted a second set of meta-analyses which included all studies examining the effects of anodal stimulation on working memory in a recently published meta-analysis (Mancuso et al., 2016). In summary, both meta-analyses did not find evidence that tDCS influences behavior on cognitive and working memory tasks.

2. P-curve analysis 1: tDCS and cognition

2.1. Methods

2.1.1. Manuscript selection criteria
Our goal for the first meta-analyses was to select empirical articles that examined the effect of tDCS on cognitive processes. First, we did a Pubmed search in 2014 of all articles with either transcranial direct current stimulation or tDCS and the following terms: language, phonological/phonology, orthographic/orthography, syntax, semantic(s), spelling, number(s), space/spatial, body schema, vision/visual, sensation, object recognition, touch, haptic(s), somatosensation/

somatosensory, attention, multisensory, decision making, decisions, learning, memory, working memory, cognitive control and cognition (569 manuscripts). We then removed any review articles, drug studies, animal studies, studies with non-healthy populations (e.g., stroke survivors, mental illness, etc.), studies that did not have a behavioral dependent variable (e.g., studies with solely subjective or neural measures were excluded), and studies that used simple motor variables (e.g., force production, finger tapping speed), as these were not considered as cognitive studies for the purpose of this analysis. We then randomized the manuscript list (223 manuscripts), and selected from this list serially for inclusion. Before starting the manuscript evaluation, we decided to stop once we had 30 articles that met inclusion criteria (Asthana et al., 2013; Balconi, Canavesio, & Vitaloni, 2014; Cattaneo, Pisoni, & Papagno, 2011; Cerruti & Schlaug, 2009; Coffman, Trumbo, & Clark, 2012; Elmer, Burkard, Renz, Meyer, & Jancke, 2009; Feeser, Prehn, Kazzer, Mungee, & Bajbouj, 2014; Ferrucci et al., 2008; Fertonani, Rosini, Cotelli, Rossini, & Miniussi, 2010; Filmer, Mattingley, & Dux, 2013; Filmer, Mattingley, Marois, & Dux, 2013; Floel et al., 2012; Fregni, Boggio, Nitsche, et al., 2005; Jacobson, Goren, Lavidor, & Levy, 2012; Kantak, Mummidisetty, & Stinear, 2012; Leite, Carvalho, Fregni, & Goncalves, 2011; Lupyan, Mirman, Hamilton, & Thompson-Schill, 2012; Marshall, Molle, Hallschmid, & Born, 2004; Marshall, Molle, Siebner, & Born, 2005; Meiron & Lavidor, 2013; Metuki, Sela, & Lavidor, 2012; Moos, Vossel, Weidner, Sparing, & Fink, 2012; Plewnia et al., 2013; Reis et al., 2015; Rosso et al., 2014; Sela, Ivry, & Lavidor, 2012; Turkeltaub et al., 2012; Vanderhasselt, De Raedt, et al., 2013; Vines, Cerruti, & Schlaug, 2008; Woods et al., 2014).

2.1.2. Selecting p-values
We followed the guidelines for p-value selection in Simonsohn et al. (2014b). In their guide, they instruct p-curve authors to generate a p-curve disclosure table (Supplemental Table 1), which contains the researcher hypothesis, study design, and the manuscript passage containing the significant test statistic. The analysis should only contain the significant test statistic(s) that test the original authors' stated hypotheses. If there was only one significant test statistic, this was included in the "results" column. If there were more than one, then the additional significant test statistics were included in the "robustness results" column. The rules for which test statistics were included in the "results" and "robustness results" columns are described in the following paragraphs. The test statistics from these columns are then used to do the p-curve analyses – a main analysis (with test statistics from the "results" column) along with a robustness analysis, as an effort to ensure that any findings in the main p-curve analysis are not idiosyncratic to those particular p-values. This information is all presented in the p-curve disclosure table (Supplemental Table 1), with information regarding tDCS parameters for each study in Supplemental Table 2.

However, a number of studies did not present test statistics or exact p-values for the hypothesis of interest or at all, or used inappropriate statistical tests. We reviewed 58 manuscripts, with 28 that did not have a p-value that could be included in the p-curve analysis (see the following section on selecting p-values). A detailed description of our selection

criteria for studies and test statistics to be included in the p-curve table, along with the relevant tables, are in the Supplemental section.

2.1.3. Computing the p-curve

After obtaining test statistics that met the stated criteria for 30 manuscripts, we used R 3.3.2 and entered the test statistics into the p-curve app (version 4.05), using R code provided at http://p-curve.com/Supplement. All code used in this paper, along with disclosure tables, scripts and outputs can be found at https://osf.io/ts6zu/. Please note that these analyses can also be done without R by using the online p-curve app (http://www.p-curve.com).

The R code first calculates the exact p-values for each hypothesis, as derived from the test statistics. From these, p-curves (plots showing the percentage of p-values within quantiles for $p < .01$, $.01 < p < .02$, ..., $.04 < p < .05$) were generated which show the distribution of these exact p-values (blue lines in the p-curve tables, see Figs. 1 and 2 for examples). Using these p-values, we then tested two hypotheses; 1. Do the studies contain evidential value (i.e., does the p-curve demonstrate right skew) and 2. Do the studies lack evidential value (i.e., is the p-curve flatter than what would be expected for a set of studies with 33% power)?

For the first hypothesis, the null hypothesis is no effect of tDCS in our sample (i.e., a flat p-curve – red dotted lines in Figs. 1 and 2), with the alternative hypothesis being that there is some effect of tDCS in our sample (a right-skewed p-curve). To do this, we first took the exact p-value from each study and generated a pp-value, which is the probability of observing this p-value if the null hypothesis is true. For example, in examining whether the p-values have evidential value (i.e., does the p-curve demonstrate right skew), this value is pp/.05. Therefore, a p-value of .01 would have a pp-value of .2, whereas a p-value of .04 would have a pp-value of .8. Given that no p-values greater than .05 are entered into the analysis, the largest pp-value would be 1. Next, the pp values were aggregated using Stouffer's method (see Simonsohn, Simmons, & Nelson, 2015). This method combines multiple pp-values, and examines whether they allow us to reject the null hypothesis of a flat p-curve (with a resulting Z-score and p-value). For example, four

studies with a flat p-curve distribution (.01, .02, .03, .04) would result in a Stouffer's z value of 0, a p-value of .5, and not reject the null hypothesis. However, four studies with p-values of .01 (pp-values of .2) would reject the null hypothesis (Z = 1.68, $p = .046$), providing evidence for a right-skewed p-curve and evidential value. Stouffer's test was used for both the full p-curve (all p values less than .05) and the half p-curve (all p values less than .025).

For the second hypothesis, is the p-curve *flatter* than what would be expected for a set of studies with 33% power, our null hypothesis is that the studies in the analysis contain 33% power (green dashed lines in the p-curve tables, Figs. 1 and 2), with the alternative hypothesis being a p-curve with significantly *less* than 33% power. The method was the same as for the first hypothesis except for the manner in which pp-values were generated. Here, power refers to the odds of obtaining a significant result given the true effect size. For example, if a researcher were to compare means of randomly generated data with $\alpha = .05$, they would obtain a significant result 5% of the time. A study with 33% power would be expected to obtain a significant result on only 1 out of 3 experiments.

Pp-values are the probability of observing a significant p-value if the null hypothesis is true. Here, the null hypothesis is a right-skewed p-curve generated from studies with 33% power. As noted in Simonsohn et al. (2014b, p. 538), studies with 33% power will generate p-values greater than .01 57.6% of the time, whereas p-values greater than .04 would be observed only 10.4%. Therefore, studies with a p-value of .01 have a pp-value of .576, and studies with a p-value of .04 have a pp-value of .104. As before, Stouffer's test was used to aggregate these p-values, both for the full p-curve and half p-curve, to examine if the selected p-values provided evidence to reject the null hypothesis (data generated from studies with 33% power).

Finally, to estimate the statistical power of the studies in the meta-analysis, the p-curve analysis compares the shape of the generated p-curve to hypothetical p-curves generated with power ranging from 5% (flat or left skewed) to 99% (right skewed – for details, see Simonsohn et al. (2015)). This analysis provides an estimate and confidence interval of the statistical power of the included studies.
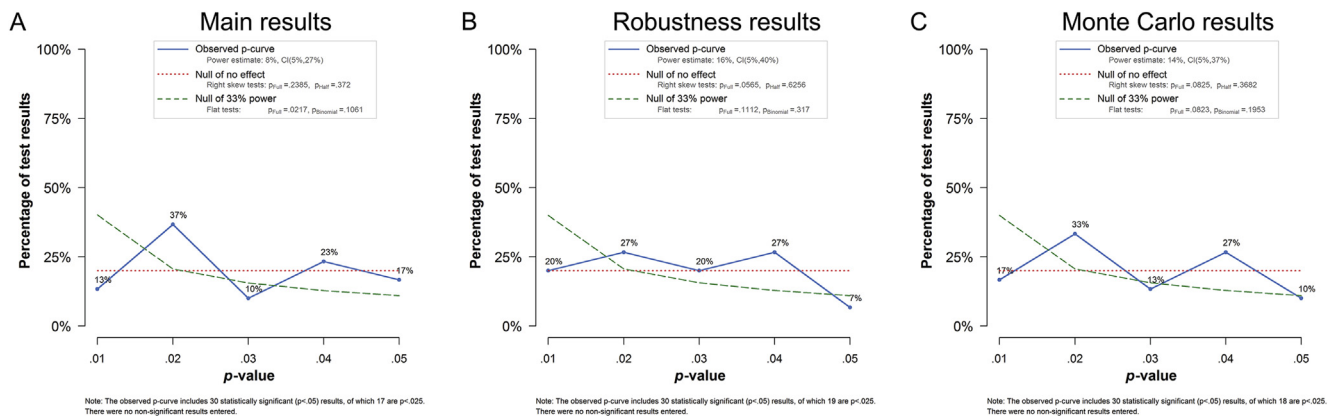

Fig. 1 – P-curves showing the distributions of *p*-values for the 30 cognitive studies (blue), the expected curve if the included studies had 33% power (green dashed line), and the expected p-curve if the included studies had no effect (red dashed). Plots are for the main results (a), robustness results (b) and Monte Carlo results (c).
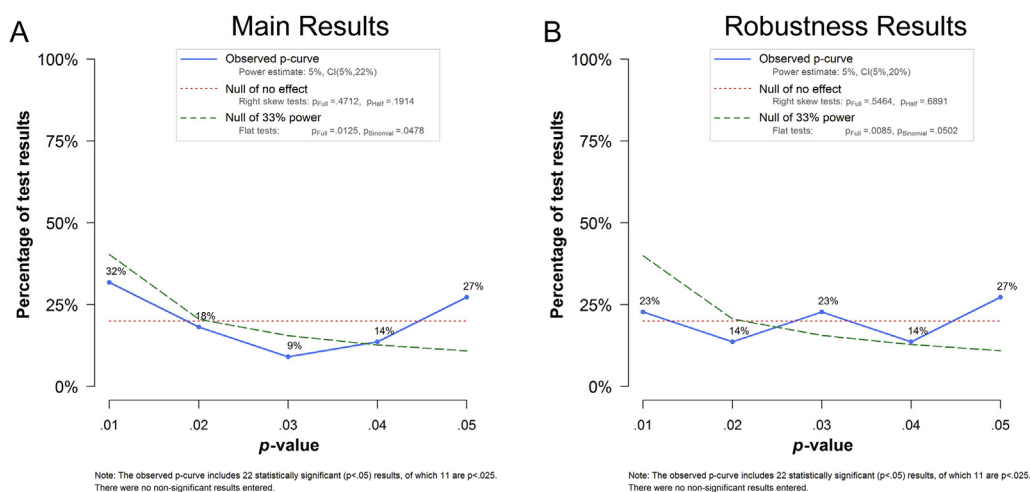
Fig. 2 — P-curves showing the distributions of *p*-values for the 22 working memory studies from the Mancuso et al. (2016) meta-analysis (blue), the expected curve if the included studies had 33% power (green dashed line), and the expected p-curve if the included studies had no effect (red dashed). Plots are for the main results (a) and robustness results (b).

2.2. Results

We entered the test statistics from the main results column of Supplemental Table 1, with the p-curve displayed in Fig. 1a. First, we examined whether the included studies contained evidential value — if the p-curve is significantly right skewed. According to Simonsohn et al. (2015), a set of studies have evidential value if either a) the half p-curve shows evidence of right skew ($p < .05$) or b) both the half and full p-curves show evidence of right skew (both with ps < .10). In our main analysis, neither criteria were met (full p-curve, $Z = -.71$, $p = .238$; half p-curve, $Z = -.33$, $p = .372$). Next, we examined whether the observed p-curve was significantly flatter when compared to an expected p-curve from studies with 33% power. This test was significant ($Z = -2.02$, $p = .022$), providing evidence that this selection of studies has inadequate (flatter than 33% power) evidential value. Finally, the estimate of the statistical power of the included studies was 8% (90% CI: 5–27%). That is, the best estimate of how many of the studies in the meta-analysis would find significant results, if replicated, is 8%.

Next, we ran the same analyses, replacing test statistics from the main results column with robustness results (10 findings). For studies with more than one result in the robustness column, we selected from these values at random. For the robustness results, this p-curve did not reject the null hypothesis that these studies demonstrate no effect (full p-curve: $Z = -1.58$, $p = .057$, half p-curve, $Z = .32$, $p = .626$; see Fig. 1b). However, we did not find evidence that the studies evidential value was inadequate (flatter than 33% power; $Z = 1.22$, $p = .111$). The estimate of statistical power from the robustness p-curve was 16% (90% CI: 5–40%).

Given that there were a number of studies (10) with multiple test statistics to choose from, we wanted to ensure that our results were not due to a random sample that was biased (e.g., more high/low *p*-values) in any particular manner. Therefore, we used a Monte Carlo sampling method to select the median set of *p*-values for inclusion in the following p-curve analysis. For each of the 30 studies, we randomly selected one *p*-value (from the listed *p*-values in the main and robustness results columns of Supplemental Table 1) from each study. This was done 10,001 times. For each of the 10,001 simulations, we calculated the average of the 30 selected *p*-values (min: .0191, max: .0249, median: .0218). We then selected the *p*-values that generated the *median* average *p*-value (average *p*-value of .0218) and then entered these into a p-curve analysis (see Fig. 1c). As before, this p-curve did not reject the null hypothesis that these studies demonstrate no effect (full p-curve: $Z = -1.39$, $p = .082$; half p-curve, $Z = -.34$, $p = .368$). However, this analysis did not demonstrate (at an alpha of .05) that the evidential value of these studies was inadequate (flatter than 33% power, $Z = -1.39$, $p = .082$). The estimate of statistical power was 14% (90% CI: 5–37%).

Finally, we examined whether our p-curve analysis was sufficiently powered to test whether the selected studies have evidential value. Using R code provided with Simmons and Simonsohn (2017), we examined the power of our p-curve analysis (see Reject_null.R on OSF). We found that with 30 studies, a p-curve analysis would have a 93.24% chance to detect evidential value (>33% power), and a 99.67% chance to detect 50% power. Furthermore, our analysis was also robust to having a number of studies with no true effect. Using the same R code, we would have an 80% chance of detecting a right-skewed p-curve if 10 studies had $d = .8$, and 20 studies had a null effect ($d = 0$). Furthermore, the power to test whether our selected studies' evidential value is inadequate is 93.64% (see Accept_null.R on OSF). To summarize, our p-curve analyses have substantially more power than the tDCS studies included in the p-curve analysis.

2.3. Discussion

Our p-curve meta-analysis examined two questions: do the studies examining the influence of tDCS on cognitive

processes contain evidential value (i.e., is the p-curve more right-skewed than a flat p-curve), and do these studies lack evidential value (is the p-curve significantly flatter than one generated from studies with 33% power). We analyzed data from three p-curves, using data from the main results table, robustness results table, and one in which we selected p-values that would be closest to the median. None were able to accept the alternative hypothesis of a right-skewed p-curve, with all failing to reject the null hypothesis of a flat p-curve. Therefore, none of these analyses showed that the *p*-values from the included studies contained evidential value. For the question of whether these studies lack evidential value (is there less than 33% power), only one of the three analysis rejected the null hypothesis. Therefore, we will not make any claims that these studies have less than 33% power. That said, the estimated power of the included tDCS studies was quite low, ranging from 8 to 16%. This means that our best estimate is that anywhere from 1 in 6 to 1 in 12 of the included studies would replicate. Finally, our power analysis demonstrates that our sample was well-powered to detect a right-skewed p-curve, even at >33% power. Overall, these results indicate minimal evidential value in the included sample of tDCS studies on cognition.

There are two potential criticisms of this meta-analysis. The cognitive studies in our sample are quite heterogeneous, with a number of different processes that may be differentially influenced by tDCS. It is possible that tDCS may have a strong influence on cognitive process A, but not cognitive process B; that tDCS may only influence cognitive processes using electrode montage X, but not electrode montage Y; or that there is some specific combination of montage, task, cognitive state, etc. that would reliably demonstrate effects. Our sample size is not large enough to examine whether tDCS influences a specific cognitive process using a specific montage, dosage, etc. However, we first note that our meta-analysis is examining whether *published manuscripts* examining the effects of tDCS on cognition have evidential value, not whether tDCS under particular circumstances does or does not influence cognitive processes. Given that all of the papers included in the p-curve analyses claim that tDCS influences cognition in some way, one would expect that tDCS would influence cognition over enough of these studies to demonstrate evidential value. That said, there is a valid concern that this analysis could be misinterpreted as definitively stating that *all* tDCS studies of cognition do not contain evidential value. To be clear, we are not making this claim.

Second, this p-curve analysis has included a sample of tDCS studies on cognition, not every tDCS study in the entire literature. There are potential concerns regarding the sample itself, such that it may be biased (either due to vagaries in random selection, or implicit/explicit biases on the part of the authors). To address both concerns about our sample and study heterogeneity, we decided to do a second meta-analysis on a full selection of studies within a specific cognitive domain. Furthermore, to ensure no selection bias on the part of the authors, we selected our experiments from a recently published meta-analysis of the effects of tDCS on a specific cognitive process — working memory.

Mancuso et al. (2016) recently published a meta-analysis of studies that aimed to improve working memory using anodal tDCS. This meta-analysis examined how four different combinations of anodal stimulation (left DLPFC, left DLPFC with cognitive training, right DPLFC, and right parietal) compared to sham stimulation influenced working memory performance. Interestingly, a number of these studies did not show a significant effect of tDCS when simply comparing performance after anodal versus sham tDCS. Out of 23 studies that compared left anodal DLPFC stimulation to sham stimulation, only five reported a significant effect of anodal tDCS compared to sham tDCS. However, 20 out of these 23 studies reported a significant effect of anodal tDCS that was consistent with their hypothesis. In most cases, the significant finding in support of the authors' hypothesis was an interaction with some other factor (e.g., item subcategories, test type, median splits, etc.). One possibility is that these additional factors are important and provide key information regarding what may (or may not) modulate the effects of tDCS on working memory. A second possibility is that having a number of statistical tests, along with increased researcher degrees of freedom (see Simmons et al., 2011) can lead to an increased number of false positives. For example, the odds of a false positive using an ANOVA with an α of .05 in a $2 \times 2 \times 2$ design with 3 DVs is $1 - (.95)^{21} = 65.9\%$. (Each ANOVA has seven tests: three main effects, three two-way interactions, and one three-way interaction; seven tests * three DVs = 21 statistical tests.) A number of potential statistical tests, combined with additional researcher degrees of freedom provided by median splits and other data transformations could lead to a literature with minimal signal. Therefore, in our second meta-analysis, we used the studies selected by Mancuso et al. (2016) in a p-curve analysis.

3. P-curve analysis 2: tDCS and working memory

3.1. Methods

The methods were the same as in Experiment 1, with a few key differences regarding study selection. Instead of selecting from a random sample of studies, we chose to examine every experiment that was analyzed in Mancuso et al. (2016). This included 34 experiments from 31 manuscripts. The p-curve disclosure table for the selected studies is in Supplemental Table 4. The Mancuso et al. (2016) meta-analysis focused on simply examining the effects of anodal versus sham tDCS on working memory (i.e., no interactions with other factors). This analysis was not significant in the majority of the experiments. Given that the p-curve analysis can only use significant findings, this precluded us from doing a p-curve analysis on the exact tests analyzed in Mancuso et al. (2016). Therefore, as done in our first p-curve analysis, we selected the significant statistical test(s) that examined the researchers' stated hypothesis. Second, if there were multiple significant findings with different polarity comparisons (e.g., anodal *vs* sham, cathodal *vs* sham), we chose the anodal *vs* sham comparison for the main results column, with the other comparison(s) in the robustness results table. If there were multiple significant

DVs from the same polarity comparison, we randomly chose which ones to include in the main versus robustness results table. Third, given that the 0-back and 1-back tasks are not considered as strong tests of working memory (see Braver et al., 1997) and were not included in the Mancuso et al. meta-analysis, we did not include significant results from these tasks in our meta-analysis. After removing 12 studies from 12 manuscripts that met our exclusion criteria, 22 studies from 19 manuscripts were included in the analyses that follow (Berryhill, Wencil, Coslett, & Olson, 2010; Bona & Silvanto, 2014; Fregni, Boggio, Nitsche, et al., 2005; Gill, Shah-Basak, & Hamilton, 2015; Gladwin, den Uyl, Fregni, & Wiers, 2012; Hoy et al., 2013; Hsu, Tseng, Liang, Cheng, & Juan, 2014; Jones & Berryhill, 2012; Jones, Goezenman, & Berryhill, 2015; Keeser et al., 2011; Keshvari, Pouretemad, & Ekhtiari, 2013; Meiron & Lavidor, 2013; Mylius et al., 2012; Nozari & Thompson-Schill, 2013; Tseng et al., 2012; Vanderhasselt, Brunoni, Loeys, Boggio, & De Raedt, 2013; Wolkenstein & Plewnia, 2013; Wu et al., 2014).

3.2. Results

Using the test statistics from the main results column of Supplemental Table 4, we generated a p-curve (see Fig. 2a) and examined if the p-curve is significantly right skewed (null of no effect, alternative: right-skewed). The null hypothesis was not rejected (full p-curve, $Z = -.07$, $p = .471$; half p-curve, $Z = -.87$, $p = .191$). We then examined whether the observed p-curve was significantly flatter when compared to an expected p-curve from studies with 33% power (null of 33% power, alternative: less than 33% power). This test was significant ($Z = -2.24$, $p = .013$), providing evidence that this selection of studies lacks (flatter than 33% power) evidential value. Finally, the estimate of statistical power of the included studies was 5% (90% CI: 5−22%). In other words, the best estimate of how many of the studies would replicate is 5%. We note that a sample of studies generated from random noise would have the same power estimate, as with an α of .05 and an effect size of zero, 5% of studies should find a significant effect.

We ran the same analysis replacing test statistics from the main results column with the robustness results column (9 findings). We did not find evidence that the studies contained evidential value (full p-curve: $Z = .12$, $p = .546$, half p-curve, $Z = .49$, $p = .689$; see Fig. 2b). Furthermore, we did find evidence that the studies evidential value was inadequate (flatter than 33% power; $Z = -2.39$, $p = .0085$). The estimate of statistical power from the robustness p-curve was also 5% (90% CI: 5−20%). Again, the best estimate of power was the same as what would be expected from studies generated from random data.

4. General discussion

Our analysis shows that there is minimal to no evidence that tDCS influences cognitive processes more generally, and working memory more specifically, in the current literature. None of our analyses provided evidence that the studies reviewed contained evidential value that tDCS influenced cognition. The estimate of statistical power of these studies

ranged from 8 to 16%, meaning that approximately 8−16% of these studies would be predicted to replicate. Given potential concerns regarding study selection and heterogeneity, we ran p-curve analyses on studies in the working memory literature from a recent meta-analysis (Mancuso et al., 2016). None of these analyses provided evidence that these studies contained evidential value, and our analyses confirmed that these studies lacked evidential value (less than 33% power). An estimate of the statistical power of these working memory studies was the same as what would be expected with random data. Although there are hundreds of published studies demonstrating that tDCS influences cognition or working memory, our meta-analyses suggests that these results should be taken with great caution. We note that our p-curve analysis differs substantially from more traditional meta-analyses. Previous tDCS meta-analyses do not correct for publication bias (Price et al., 2015) whereas others use methods like Trim-and-Fill that underestimate the influence of publication bias. Our study demonstrates the importance of taking publication bias into account in examining evidential value in a set of studies.

Do these results provide definitive evidence that tDCS *does not* influence cognitive processes? No. First, we note that the studies examined in this review utilizes a variety of different stimulation parameters, in which stimulation site, polarity, current, reference electrode location, length of stimulation, when stimulation occurred (during or after the task), and other factors varied. One possibility is that there are specific tDCS combinations that consistently influence performance in a specific cognitive domain. Consider a hypothetical example in which anodal tDCS only has an effect on task A at 2.0 mA, but no effect on task A at 1.5 mA, and no effect on task B regardless of amperage. A p-curve analysis that includes all of these studies may not find a right-skewed p-curve, especially if the effect size of anodal tDCS on task A at 2.0 mA is small to moderate. It could be that a few of the studies in our p-curve analysis have a true, replicable effect, while the rest are noise masquerading as signal.

Regardless, even if there is evidential value in some of the included studies, the overall p-curve analysis suggests that the majority of the included studies have limited to no evidential value. To clarify, we note that studies included in a p-curve analysis do not need to have particularly large, or uniform, effect sizes to generate a right-skewed p-curve (see Simonsohn et al., 2014b, Fig. 1B which shows a right-skewed p-curve generated from data with $N = 20$ and $d = .3$, a relatively small sample size and effect size). Furthermore, our cognition p-curve analysis was sufficiently powered to detect a right-skewed p-curve if one-third of the studies had a large effect size ($d = .8$) and two-thirds of the studies in the sample had no effect ($d = 0$). Therefore, it is very likely that a number of studies examining the effects of tDCS on working memory and cognition in our sample have minimal to no evidential value. Furthermore, given typical assumptions regarding random sampling (i.e., that the random sample is representative of its population), it is likely that there are a number of published tDCS studies on cognition and working memory that have minimal to no evidential value. This is a significant problem for the field. A number of researchers may be wasting time attempting to replicate findings that are not true.

Furthermore, if there is a true effect of tDCS given a particular combination of task, montage, polarity, etc., these real findings run the risk of being hidden in a sea of noise. To address these concerns, we outline a few reasons for why these studies may have minimal evidential value, and provide suggestions for future research.

One possibility is the sample sizes typically seen in the tDCS literature. In the studies examined in this paper, the average N for each group in between-subjects designs was 14.6 (9.5−21), and 17.9 (10−46) for within-subjects designs. We used G-Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) to examine the number of subjects needed to detect an effect with 80% power (i.e., there being an 80% chance to detect a significant effect if an effect existed) using a simple two-tailed t-test. For between-subjects designs, one needs a n of 26 in each group to detect a large effect ($d = .8$), 64 in each group to detect a medium effect ($d = .5$), and 394 to detect a small effect. None of the between-subjects studies in our analysis had enough subjects to detect a large effect with 80% power, with 15 subjects per group only having 69% power to detect a large effect (38% medium effect, 13% small effect). To have 80% power in a within-subjects design, one would need 15 participants with a large effect size, 34 for a medium effect size, and 199 for a small effect size. The average within-subjects tDCS study in this review ($n = 18$) did have sufficient power to detect a large effect (89%), but only 51% power to detect a medium-sized effect, and 13% to detect a small effect. Furthermore, we note that these power estimates are only for simple comparisons between two groups. Given that many of these studies use more complicated designs (e.g., 3×2 or $3 \times 2 \times 2$ interactions), it is possible that more participants would be necessary in these designs. Future tDCS studies should have larger sample sizes to provide better estimates of the actual effect (if any) of tDCS on cognitive processes, and to avoid false positives with falsely inflated effect sizes that occur with underpowered studies. We note that relatively small sample sizes can be justified if the *true effect* is large. Robust findings in the psychological literature (e.g., the Stroop effect, Simon effect) can be easily replicated without needing to test 200 subjects. If an initial study finds a large effect of tDCS on a cognitive process, it should be relatively easy to replicate if there truly is a large effect. However, given that underpowered studies can only detect large effects (the "winner's curse", Button et al., 2013; Ioannidis, 2008), it is imperative that these replications be done to examine whether published findings have evidential value.

Therefore, we encourage direct replication and publication (even of null results) of findings in the tDCS literature. For example, the study with the lowest p-value across both of our p-curve analyses (Cattaneo et al., 2011, $p = .0001$) recently failed to replicate with $p = .97$ (Vannorsdall et al., 2016). Furthermore, there have been other recent failures to replicate tDCS findings in the literature (e.g., Brückner & Kammer, 2016; Nilsson, Lebedev, Rydström, & Lövdén, 2017; Westwood et al., 2017). There is a well-known bias towards publishing positive results (especially if they are novel and interesting) versus null results. This weighting of novelty over veracity and replicability, along with limitations in journal pages and interest, can make it difficult to published failed findings.

Given these potential biases, it is absolutely critical for researchers to be aware of what can and cannot replicate. Therefore, we suggest that researchers open their file drawers, either by attempting to publish these findings (in traditional journals or preprint archives) or posting the data and findings on sites like the Open Science Framework (https://osf.io/). Furthermore, pressures to publish − especially evident for graduate students, post-doctoral fellows, and early-career faculty whose career success is often predicated on published papers − provide perverse incentives for scientists. We also suggest that editors and funding agencies reward solid, replicable research over unstable findings.

Furthermore, the results from this p-curve analysis provide evidence that tDCS studies on cognition and working memory should be further examined. Although all 52 studies were published with significant results demonstrating tDCS influencing cognitive processes, in sum, the results provide minimal evidence in support of this claim. Given this, we present a few recommendations for future tDCS studies, such that the evidential value of tDCS can be better examined. First, tDCS studies should be preregistered, with the hypothesis, dependent variables, number of subjects tested (using power analyses to choose sample size) and analysis pipeline described before the data are collected. In our sample of tDCS studies, a number of studies had three factors and multiple dependent variables. These designs alone, with no a priori hypothesis regarding the specific hypothesis of interest, allow for high false positives. By preregistering the important hypotheses to be tested, along with sample sizes that are designed to be sufficiently powered to detect the finding of interest, the reader can be more confident that a significant effect indicates that tDCS influences cognition. Furthermore, preregistration limits a number of questionable research practices (e.g., HARKing, adding subjects until reaching significance, etc.) that decrease the evidential value of a statistically significant finding. Other suggestions for improving study design, both for researchers and reviewers, can be found in Simmons et al. (2011).

Finally, we note that these issues are not limited to studies of transcranial direct current stimulation. Although this manuscript focused on tDCS and cognition/working memory, we note that other subfields in psychology have been struggling with similar issues regarding replicability and evidential value (Chambers, 2017). It is imperative that researchers across all fields endeavor to use better methods so that future work can build on a firm foundation (Kaelin, 2017). Given the potential for tDCS in cognitive enhancement, both for neurologically intact and brain-damaged individuals, it is important that we have a clear understanding of its benefits. Therefore, we encourage the field to use research practices that allow for a better understanding of the evidential value of tDCS.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2017.06.021.

REFERENCES

Asthana, M., Nueckel, K., Muhlberger, A., Neueder, D., Polak, T., Domschke, K., et al. (2013). Effects of transcranial direct current stimulation on consolidation of fear memory. *Front Psychiatry, 4*, 107. http://dx.doi.org/10.3389/fpsyt.2013.00107.

Balconi, M., Canavesio, Y., & Vitaloni, S. (2014). Activation of the prefrontal cortex and posterior parietal cortex increases the recognition of semantic violations in action representation. *Brain Stimulation, 7*(3), 435—442. http://dx.doi.org/10.1016/j.brs.2014.01.011.

Banissy, M. J., & Muggleton, N. G. (2013). Transcranial direct current stimulation in sports training: Potential approaches. *Front Hum Neurosci, 7*, 129. http://dx.doi.org/10.3389/fnhum.2013.00129.

Berryhill, M. E., Wencil, E. B., Coslett, H. B., & Olson, I. R. (2010). A selective working memory impairment after transcranial direct current stimulation to the right parietal lobe. *Neuroscience Letters, 479*(3), 312—316. http://dx.doi.org/10.1016/j.neulet.2010.05.087.

Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., et al. (2016). Safety of transcranial direct current stimulation: Evidence based update 2016. *Brain Stimulation, 9*(5), 641—661. http://dx.doi.org/10.1016/j.brs.2016.06.004.

Bindman, L. J., Lippold, O. C. J., & Redfearn, J. W. (1962). Long-lasting changes in the level of the electrical activity of the cerebral cortex produced by polarizing currents. *Nature, 196*(4854), 584—585. http://dx.doi.org/10.1038/196584a0.

Boggio, P. S., Rigonatti, S. P., Ribeiro, R. B., Myczkowski, M. L., Nitsche, M. A., Pascual-Leone, A., et al. (2008). A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. *Int J Neuropsychopharmacol, 11*(2), 249—254. http://dx.doi.org/10.1017/S1461145707007833.

Boggio, P. S., Zaghi, S., & Fregni, F. (2009). Modulation of emotions associated with images of human pain using anodal transcranial direct current stimulation (tDCS). *Neuropsychologia, 47*(1), 212—217. http://dx.doi.org/10.1016/j.neuropsychologia.2008.07.022.

Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: Behavioral and neural dissociations. *Plos One, 9*(3). http://dx.doi.org/10.1371/journal.pone.0090808.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage, 5*(1), 49—62. http://dx.doi.org/10.1006/nimg.1996.0247.

Brückner, S., & Kammer, T. (2016). No modulation of visual cortex excitability by transcranial direct current stimulation. *Plos One, 11*(12), e0167697.

Brunoni, A. R., & Vanderhasselt, M. A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: A systematic review and meta-analysis. *Brain and Cognition, 86*, 1—9. http://dx.doi.org/10.1016/j.bandc.2014.01.008.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience, 14*(5), 365—376. http://dx.doi.org/10.1038/nrn3475.

Cattaneo, Z., Pisoni, A., & Papagno, C. (2011). Transcranial direct current stimulation over Broca's region improves phonemic and semantic fluency in healthy individuals. *Neuroscience, 183*, 64—70. http://dx.doi.org/10.1016/j.neuroscience.2011.03.058.

Cerruti, C., & Schlaug, G. (2009). Anodal transcranial direct current stimulation of the prefrontal cortex enhances complex verbal associative thought. *Journal of Cognitive Neuroscience, 21*(10), 1980—1987. http://dx.doi.org/10.1162/jocn.2008.21143.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice.* Princeton University Press.

Coffman, B. A., Trumbo, M. C., & Clark, V. P. (2012). Enhancement of object detection with transcranial direct current stimulation is associated with increased attention. *BMC Neuroscience, 13*. http://dx.doi.org/10.1186/1471-2202-13-108.

Dedoncker, J., Brunoni, A. R., Baeken, C., & Vanderhasselt, M. A. (2016). A systematic review and meta-analysis of the effects of transcranial direct current stimulation (tDCS) over the dorsolateral prefrontal cortex in healthy and neuropsychiatric Samples: Influence of stimulation parameters. *Brain Stimulation, 9*(4), 501—517. http://dx.doi.org/10.1016/j.brs.2016.04.006.

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89—98. http://dx.doi.org/10.2307/2669529.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455—463.

Elmer, S., Burkard, M., Renz, B., Meyer, M., & Jancke, L. (2009). Direct current induced short-term modulation of the left dorsolateral prefrontal cortex while learning auditory presented nouns. *Behavioral and Brain Functions, 5*. http://dx.doi.org/10.1186/1744-9081-5-29.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175—191.

Feeser, M., Prehn, K., Kazzer, P., Mungee, A., & Bajbouj, M. (2014). Transcranial direct current stimulation enhances cognitive control during emotion regulation. *Brain Stimulation, 7*(1), 105—112. http://dx.doi.org/10.1016/j.brs.2013.08.006.

Ferrucci, R., Marceglia, S., Vergari, M., Cogiamanian, F., Mrakic-Sposta, S., Mameli, F., et al. (2008). Cerebellar transcranial direct current stimulation impairs the practice-dependent proficiency increase in working memory. *Journal of Cognitive Neuroscience, 20*(9), 1687—1697. http://dx.doi.org/10.1162/jocn.2008.20112.

Fertonani, A., Rosini, S., Cotelli, M., Rossini, P. M., & Miniussi, C. (2010). Naming facilitation induced by transcranial direct current stimulation. *Behavioural Brain Research, 208*(2), 311—318. http://dx.doi.org/10.1016/j.bbr.2009.10.030.

Filmer, H. L., Mattingley, J. B., & Dux, P. E. (2013). Improved multitasking following prefrontal tDCS. *Cortex, 49*(10), 2845—2852. http://dx.doi.org/10.1016/j.cortex.2013.08.015.

Filmer, H. L., Mattingley, J. B., Marois, R., & Dux, P. E. (2013). Disrupting prefrontal cortex prevents performance gains from sensory-motor training. *Journal of Neuroscience, 33*(47), 18654—18660. http://dx.doi.org/10.1523/jneurosci.2019-13.2013.

Floel, A., Suttorp, W., Kohl, O., Kurten, J., Lohmann, H., Breitenstein, C., et al. (2012). Non-invasive brain stimulation improves object-location learning in the elderly. *Neurobiology of Aging, 33*(8), 1682—1689. http://dx.doi.org/10.1016/j.neurobiolaging.2011.05.007.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502—1505.

Fregni, F., Boggio, P. S., Mansur, C. G., Wagner, T., Ferreira, M. J., Lima, M. C., et al. (2005). Transcranial direct current stimulation of the unaffected hemisphere in stroke patients. *NeuroReport, 16*(14), 1551—1555.

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., et al. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research, 166*(1), 23—30. http://dx.doi.org/10.1007/s00221-005-2334-6.

Gelman, A., & Carlin, J. (2014). Beyond power Calculations: Assessing type S (sign) and type M (Magnitude) errors. *Perspect Psychol Sci, 9*(6), 641—651. http://dx.doi.org/10.1177/1745691614551642.

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time* (pp. 460—465). Department of Statistics, Columbia University.

Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460.

Gill, J., Shah-Basak, P. P., & Hamilton, R. (2015). It's the thought that Counts: examining the task-dependent effects of transcranial direct current stimulation on executive function. *Brain Stimulation, 8*(2), 253—259. http://dx.doi.org/10.1016/j.brs.2014.10.018.

Gladwin, T. E., den Uyl, T. E., Fregni, F. F., & Wiers, R. W. (2012). Enhancement of selective attention by tDCS: Interaction with interference in a Sternberg task. *Neuroscience Letters, 512*(1), 33—37. http://dx.doi.org/10.1016/j.neulet.2012.01.056.

Hill, A. T., Fitzgerald, P. B., & Hoy, K. E. (2016). Effects of anodal transcranial direct current stimulation on working memory: A systematic review and meta-analysis of findings from healthy and neuropsychiatric populations. *Brain Stimulation, 9*(2), 197—208. http://dx.doi.org/10.1016/j.brs.2015.10.006.

Horvath, J. C., Forte, J. D., & Carter, O. (2015). Quantitative review finds No evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimulation, 8*(3), 535—550. http://dx.doi.org/10.1016/j.brs.2015.01.400.

Hoy, K. E., Emonson, M. R. L., Arnold, S. L., Thomson, R. H., Daskalakis, Z. J., & Fitzgerald, P. B. (2013). Testing the limits: Investigating the effect of tDCS dose on working memory enhancement in healthy controls. *Neuropsychologia, 51*(9), 1777—1784. http://dx.doi.org/10.1016/j.neuropsychologia.2013.05.018.

Hsu, T.-Y., Tseng, P., Liang, W.-K., Cheng, S.-K., & Juan, C.-H. (2014). Transcranial direct current stimulation over right posterior parietal cortex changes prestimulus alpha oscillation in visual short-term memory task. *NeuroImage, 98*, 306—313. http://dx.doi.org/10.1016/j.neuroimage.2014.04.069.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640—648. http://dx.doi.org/10.1097/EDE.0b013e31818131e7.

Jacobson, L., Goren, N., Lavidor, M., & Levy, D. A. (2012). Oppositional transcranial direct current stimulation (tDCS) of parietal substrates of attention during encoding modulates episodic memory. *Brain Research, 1439*, 66—72. http://dx.doi.org/10.1016/j.brainres.2011.12.036.

Jones, K. T., & Berryhill, M. E. (2012). Parietal contributions to visual working memory depend on task difficulty. *Frontiers in Psychiatry, 3*, 81. http://dx.doi.org/10.3389/fpsyt.2012.00081.

Jones, K. T., Goezenman, F., & Berryhill, M. E. (2015). The strategy and motivational influences on the beneficial effect of neurostimulation: A tDCS and fNIRS study. *NeuroImage, 105*, 238—247. http://dx.doi.org/10.1016/j.neuroimage.2014.11.012.

Kaelin, W. G., Jr. (2017). Publish houses of brick, not mansions of straw. *Nature, 545*(7655), 387.

Kantak, S. S., Mummidisetty, C. K., & Stinear, J. W. (2012). Primary motor and premotor cortex in implicit sequence learning — evidence for competition between implicit and explicit human motor memory systems. *European Journal of Neuroscience, 36*(5), 2710—2715. http://dx.doi.org/10.1111/j.1460-9568.2012.08175.x.

Keeser, D., Padberg, F., Reisinger, E., Pogarell, O., Kirsch, V., Palm, U., et al. (2011). Prefrontal direct current stimulation modulates resting EEG and event-related potentials in healthy subjects: A standardized low resolution tomography (sLORETA) study. *NeuroImage, 55*(2), 644—657. http://dx.doi.org/10.1016/j.neuroimage.2010.12.004.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 2*(3), 196—217. http://dx.doi.org/10.1207/s15327957pspr0203_4.

Keshvari, F., Pouretemad, H. R., & Ekhtiari, H. (2013). The polarity-dependent effects of the bilateral brain stimulation on working memory. *Basic Clin Neurosci, 4*(3), 224—231.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480—498.

Leite, J., Carvalho, S., Fregni, F., & Goncalves, O. F. (2011). Task-specific effects of tDCS-induced cortical excitability changes on cognitive and motor sequence set shifting performance. *Plos One, 6*(9). http://dx.doi.org/10.1371/journal.pone.0024140.

Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition, 124*(1), 36—49. http://dx.doi.org/10.1016/j.cognition.2012.04.002.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language, 91*, 5—27.

Mancuso, L. E., Ilieva, I. P., Hamilton, R. H., & Farah, M. J. (2016). Does transcranial direct current stimulation improve healthy working Memory?: a meta-analytic review. *Journal of Cognitive Neuroscience*. http://dx.doi.org/10.1162/jocn_a_00956.

Marshall, L., Molle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *Journal of Neuroscience, 24*(44), 9985—9992. http://dx.doi.org/10.1523/jneurosci.2725-04.2004.

Marshall, L., Molle, M., Siebner, H. R., & Born, J. (2005). Bifrontal transcranial direct current stimulation slows reaction time in a working memory task. *BMC Neuroscience, 6*. http://dx.doi.org/10.1186/1471-2202-6-23.

Meiron, O., & Lavidor, M. (2013). Unilateral prefrontal direct current stimulation effects are modulated by working memory load and gender. *Brain Stimulation, 6*(3), 440—447. http://dx.doi.org/10.1016/j.brs.2012.05.014.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512—534.

Metuki, N., Sela, T., & Lavidor, M. (2012). Enhancing cognitive control components of insight problems solving by anodal tDCS of the left dorsolateral prefrontal cortex. *Brain Stimulation, 5*(2), 110—115. http://dx.doi.org/10.1016/j.brs.2012.03.002.

Moos, K., Vossel, S., Weidner, R., Sparing, R., & Fink, G. R. (2012). Modulation of top-down control of visual attention by cathodal tDCS over right IPS. *Journal of Neuroscience, 32*(46), 16360—16368. http://dx.doi.org/10.1523/jneurosci.6233-11.2012.

Mylius, V., Jung, M., Menzler, K., Haag, A., Khader, P. H., Oertel, W. H., et al. (2012). Effects of transcranial direct current stimulation on pain perception and working memory.

European Journal of Pain, 16(7), 974–982. http://dx.doi.org/10.1002/j.1532-2149.2011.00105.x.

Nilsson, J., Lebedev, A. V., Rydström, A., & Lövdén, M. (2017). Direct-current stimulation does little to improve the outcome of working memory training in older adults. *Psychological Science, 28*(7), 907–920. http://dx.doi.org/10.1177/0956797617698139.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J Physiol, 527 Pt 3*(3), 633–639.

Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., et al. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *Journal of Cognitive Neuroscience, 15*(4), 619–626. http://dx.doi.org/10.1162/089892903321662994.

Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt? *Neuropsychologia, 51*(13), 2770–2780. http://dx.doi.org/10.1016/j.neuropsychologia.2013.08.019.

Plewnia, C., Zwissler, B., Langst, I., Maurer, B., Giel, K., & Kruger, R. (2013). Effects of transcranial direct current stimulation (tDCS) on executive functions: Influence of COMT Val/Met polymorphism. *Cortex, 49*(7), 1801–1807. http://dx.doi.org/10.1016/j.cortex.2012.11.002.

Price, A. R., & Hamilton, R. H. (2015). A Re-evaluation of the cognitive effects from single-session transcranial direct current stimulation. *Brain Stimulation, 8*(3), 663–665. http://dx.doi.org/10.1016/j.brs.2015.03.007.

Price, A. R., McAdams, H., Grossman, M., & Hamilton, R. H. (2015). A meta-analysis of transcranial direct current stimulation studies examining the reliability of effects on language measures. *Brain Stimulation, 8*(6), 1093–1100. http://dx.doi.org/10.1016/j.brs.2015.06.013.

Reis, J., Fischer, J. T., Prichard, G., Weiller, C., Cohen, L. G., & Fritsch, B. (2015). Time- but not sleep-dependent consolidation of tDCS-enhanced visuomotor skills. *Cerebral Cortex, 25*(1), 109–117. http://dx.doi.org/10.1093/cercor/bht208.

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

Rosso, C., Valabregue, R., Arbizu, C., Ferrieux, S., Vargas, P., Humbert, F., et al. (2014). Connectivity between right inferior frontal gyrus and supplementary motor area predicts after-effects of right frontal cathodal tDCS on picture naming speed. *Brain Stimulation, 7*(1), 122–129. http://dx.doi.org/10.1016/j.brs.2013.08.007.

Sela, T., Ivry, R. B., & Lavidor, M. (2012). Prefrontal control during a semantic decision task that involves idiom comprehension: A transcranial direct current stimulation study. *Neuropsychologia, 50*(9), 2271–2280. http://dx.doi.org/10.1016/j.neuropsychologia.2012.05.031.

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. http://dx.doi.org/10.1177/0956797611417632.

Simmons, J. P., & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science, 28*(5), 687–693. http://dx.doi.org/10.1177/0956797616658563.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). p-Curve and effect Size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. http://dx.doi.org/10.1177/1745691614553988.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: A key to the file-drawer. *Journal of Experimental Psychology-General, 143*(2), 534–547. http://dx.doi.org/10.1037/a0033242.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology-General, 144*(6), 1146–1152. http://dx.doi.org/10.1037/xge0000104.

Sparing, R., Dafotakis, M., Meister, I. G., Thirugnanasambandam, N., & Fink, G. R. (2008). Enhancing language performance with non-invasive brain stimulation—a transcranial direct current stimulation study in healthy humans. *Neuropsychologia, 46*(1), 261–268. http://dx.doi.org/10.1016/j.neuropsychologia.2007.07.009.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *The Neuroscientist, 17*(1), 37–53. http://dx.doi.org/10.1177/1073858410386614.

Tseng, P., Hsu, T.-Y., Chang, C.-F., Tzeng, O. J. L., Hung, D. L., Muggleton, N. G., et al. (2012). Unleashing Potential: Transcranial direct current stimulation over the right posterior parietal cortex improves change detection in low-performing individuals. *Journal of Neuroscience, 32*(31), 10554–10561. http://dx.doi.org/10.1523/jneurosci.0362-12.2012.

Turkeltaub, P. E., Benson, J., Hamilton, R. H., Datta, A., Bikson, M., & Coslett, H. B. (2012). Left lateralizing transcranial direct current stimulation improves reading efficiency. *Brain Stimulation, 5*(3), 201–207. http://dx.doi.org/10.1016/j.brs.2011.04.002.

Vanderhasselt, M. A., Brunoni, A. R., Loeys, T., Boggio, P. S., & De Raedt, R. (2013). Nosce te ipsum—Socrates revisited? Controlling momentary ruminative self-referent thoughts by neuromodulation of emotional working memory. *Neuropsychologia, 51*(13), 2581–2589. http://dx.doi.org/10.1016/j.neuropsychologia.2013.08.011.

Vanderhasselt, M. A., De Raedt, R., Brunoni, A. R., Campanha, C., Baeken, C., Remue, J., et al. (2013). tDCS over the left prefrontal cortex enhances cognitive control for positive affective stimuli. *Plos One, 8*(5). http://dx.doi.org/10.1371/journal.pone.0062219.

Vannorsdall, T. D., van Steenburgh, J. J., Schretlen, D. J., Jayatillake, R., Skolasky, R. L., & Gordon, B. (2016). Reproducibility of tDCS results in a randomized Trial: Failure to replicate findings of tDCS-induced enhancement of verbal fluency. *Cognitive and Behavioral Neurology, 29*(1), 11–17. http://dx.doi.org/10.1097/WNN.0000000000000086.

Vines, B. W., Cerruti, C., & Schlaug, G. (2008). Dual-hemisphere tDCS facilitates greater improvements for healthy subjects' non-dominant hand compared to uni-hemisphere stimulation. *BMC Neuroscience, 9.* http://dx.doi.org/10.1186/1471-2202-9-103.

Westwood, S. J., Olson, A., Miall, R. C., Nappo, R., & Romani, C. (2017). Limits to tDCS effects in language: Failures to modulate word production in healthy participants with frontal or temporal tDCS. *Cortex, 86*, 64–82.

Wolkenstein, L., & Plewnia, C. (2013). Amelioration of cognitive control in depression by transcranial direct current stimulation. *Biological Psychiatry, 73*(7), 646–651. http://dx.doi.org/10.1016/j.biopsych.2012.10.010.

Woods, A. J., Hamilton, R. H., Kranjec, A., Minhaus, P., Bikson, M., Yu, J., et al. (2014). Space, time, and causality in the human brain. *NeuroImage, 92*, 285–297. http://dx.doi.org/10.1016/j.neuroimage.2014.02.015.

Wu, Y.-J., Tseng, P., Chang, C.-F., Pai, M.-C., Hsu, K.-S., Lin, C.-C., et al. (2014). Modulating the interference effect on spatial working memory by applying transcranial direct current stimulation over the right dorsolateral prefrontal cortex. *Brain and Cognition, 91*, 87–94. http://dx.doi.org/10.1016/j.bandc.2014.09.002.