A systematic investigation into the effects of anodal tDCS on healthy populations across measures of language, working memory & novel language acquisition

SAMUEL WESTWOOD

Doctor of Philosophy

ASTON UNIVERSITY

September, 2017

THESIS SUMMARY

Transcranial direct current stimulation (tDCS) is a technique thought to modify cognition via a weak electric current applied to the scalp. Several thousand papers have been published since its inception in the early 2000s, with positive effects observed across healthy and patient samples in terms of language, memory, attention and various other executive functions. However, evidence is emerging that reported effects are exaggerated, and difficult to reproduce, especially in studies using single applications of anodal tDCS on healthy individuals. This thesis documents several studies that aimed to verify whether tDCS can modify word production, working memory and novel language acquisition in healthy participants when using conventional stimulation parameters, whilst considering factors that have driven its unreliability.

The peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.

~ John Stuart Mill

ACKNOWLEDGMENTS

This PhD would not have been possible without the patience and unyielding belief in me from my supervisor, Dr Cristina Romani. Thanks also to my co-supervisors, Prof Chris Miall and Dr Andrew Olson, with special thanks to Prof Chris Miall for the ludicrous suggestion that I do a PhD in the first place!

Thank you Raffaele for our many conversations about science and – of course – Breaking Bad. These kept me sane!

Thank you Jacqueline Benfield for helping with data collection in Experiment 2 reported in Chapter 2.

To Georgina Heyes, a dear friend who gave me confidence when I needed it most. I am sorry.

I am forever indebted to Alena Nash. If it were not for you, this thesis would remain unwritten. Thank you for your kindness, for your support, and love. Спасибо.

Finally, I want to thank my mother, Kim Westwood, for my sheer bloody mindedness, which she in turn inherited from her mother, Marylin Hancox, to whom I am also indebted and to whom I dedicate this PhD.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

Chapter 1: Introduction

Recent years have witnessed a growth in efforts to enhance cognition in the healthy and the pathological brain. Intensive brain training (Kable et al., 2017), pharmaceuticals (Dresler et al., 2017), meditation (Harris, 2014), recreational drugs (Hitchens, 2016), spiritual enlightenment (Kohls, Sauer, Offenbächer, & Giordano, 2011), are just a few examples. Particular attention, however, is being paid to non-invasive brain stimulation techniques that administer a very weak electric current via two electrodes placed on the scalp. The most popular form of this technique is transcranial direct current stimulation (or tDCS). Following pioneering work in the early 2000s that showed tDCS can modulate motor cortical excitably, researchers sought to extend these findings to the cognitive domain. The result was several thousand reports of improved language, memory, attention, decision-making, learning, prejudice, affect, and risk-taking, and many other core cognitive functions (see Santarnecchi et al., 2015). However, recent evidence is now pointing out that findings are exaggerated; difficult to reproduce, particularly in single applications in healthy participants; and based on sample sizes with low evidential value. Thus, the picture is confusing; a state of affairs that cannot continue given the high impact tDCS poses for applications both within healthy and patient populations. This thesis is a contribution to this question because it systematically investigates whether tDCS can modify word production, working memory and novel language acquisition when applied to healthy participants with widely used, conventional stimulation protocols. I focused on these areas of cognition because they are popular areas of study in the tDCS literature, and because they give me scope to consider reasons for both the unreliability of outcomes and the disparity between evidence for and against tDCS efficacy. Before I move on to discussing the work I have undertaken, I will spend the rest of this introductory chapter reviewing literature pertinent to this thesis then outline its aims and a summary of the work reported in each chapter.

1.1 transcranial Direct Current Stimulation (or tDCS)

Transcranial direct current stimulation (or tDCS) is thought to modulate cortical excitability via a weak current passed via scalp electrodes. The device consists of 1) two

electrodes (one anode; one cathode) insulated in non-conductive rubber; 2) two sponges, into which each electrode is inserted; 3) two rubber insulated connecting wires; 4) a contact medium (i.e., salty water or saline solution) into which the sponges are dipped to reduce current impedance as the current passes through the scalp; 4) a 9v battery with an adjustable current; and 5) two non-conductive rubber straps. The electrodes vary in both shape and size, although the most common is a square electrode 9 to 100cm^2 in size. In a typical setup, electrodes are fastened to the scalp with the rubber straps: one electrode placed over the brain area of interest, whilst the other electrode is placed on the non-relevant area, such as the forehead, cheek or arm. The current is thought to flow through the path of least resistance from the positive anode to the negative cathode, with around 50 percent reaching underlying cortical tissue thanks to current shunting – i.e., deflection of the current through conductive tissue, such as skin, fat, skull bone and cerebral spinal fluid (see Nitsche et al., 2008; Nitsche & Paulus, 2000; Reinhart, Cosman, Fukuda, & Woodman, 2017; Rush & Driscoll, 1968; but, see also Underwood, 2016; Vöröslakos et al., 2018).

1.2 Proposed Mechanisms of Action in Humans

It is widely thought that directional changes in excitability occur under each electrode, with an increase in excitation underneath the anode (known as anodal tDCS) and a decrease in excitation underneath the cathode (known as cathodal tDCS). The physiological basis for these excitability changes in humans still remains unknown, but research has begun to elucidate possible mechanisms of action. The main source of this information is from pharmacological studies applying tDCS with the "classic" montage (i.e., placing 35cm^2 electrodes over the primary motor cortex and the contralateral supraorbital area) whilst administering drugs to modulate the activity of neurotransmitters. These studies reveal that the effects of tDCS evolve overtime, and are underpinned by several neurochemical and neurophysiological changes (for extensive reviews, see Giordano et al., 2017; Pelletier & Cicchetti, 2015; Stagg & Nitsche, 2011).

The primary mechanism of action is the transient modulation of trans-membrane potentials in the first few seconds or minutes of stimulation. These changes do not cause effects that persist post stimulation cessation, and are driven by the modification of ion channel activity.

For example, voltage-dependent sodium channel blocker carbamazepine (CBZ) reduces whilst the calcium channel blocker flunarizine (FLZ) abolishes the effects of anodal tDCS. Alternatively, neither CBZ nor FLZ can modify excitability shifts induced by cathodal tDCS, which is to be expected since cathodal induced hyperpolarization leads to inactivation of the relevant ion channels, rendering ion channel blockers ineffective (see Nitsche et al., 2003). Moreover, the acute effects of tDCS appear to not depend on glutamate or GABA – two of the most important excitatory and inhibitory neurotransmitters in the brain that underpin neural plasticity. Neither dextromethorphane (DMO), an N-methyl-D- aspartate (or NMDA) receptor (a subtype of glutamate receptor) antagonist, nor lorazepam (LOR), a GABA receptor agonist, modify the effect of tDCS for either polarity when applied for a four seconds (a period too short to induce after effects), suggesting that at least the acute effects are largely underpinned by modification of ion channel activity rather than neuroplasticity (see Nitsche et al., 2003; Nitsche et al., 2004a).

Over time (~5mins), the initial polarization described in the previous paragraph leads to neuroplasticity changes, which prolong the effects seen during tDCS up to one-hour post stimulation cessation. For example, ion channel blockers FLZ and CBZ can abolish the after-effects of both tDCS polarities (Liebetanz, Nitsche, Tergau & Paulus, 2002; Nitsche et al., 2003), whilst administering LOR significantly reduces excitability in the first 10 minutes post anodal tDCS (Nitsche et al., 2004). Cortical excitability changes following anodal and cathodal tDCS are wiped out with DMO (Liebetanz et al., 2002; Nitsche et al., 2003), whereas d-cycloserine (CYC), an NMDA receptor agonist, prolongs the duration of anodal tDCS aftereffects (Nitsche et al., 2004b). Similarly, studies using magnetic resonance spectroscopy (MRS), a technique which measures concentrations of neurochemicals within certain regions of interest in the brain, show a respective increase versus decrease in concentrations of glutamate versus GABA after anodal tDCS and a reduction in glutamate after cathodal tDCS (Clark, Coffman, Trumbo & Gasparovic, 2011; Stagg et al., 2009, Stagg, Bachtiar, & Johansen-Berg, 2011)

1.3 Parameters

The parameters in which to apply tDCS were investigated mainly by studies using the "classic" montage, with 1mA of current usually applied for up to roughly 20 minutes (Nitsche & Paulus, 2000; Nitsche et al., 2008). Key parameters that can influence the neurophysiological impact of tDCS include the stimulation intensity, duration and electrode montage.

1.3.1 Intensity

The efficacy of tDCS relies on the current density, which is the current intensity relative to the size of the electrode surface. Modulation of cortical excitability has been achieved in current densities between 0.03 – 0.08 mA/cm^2, with typical current intensities set to 1 to 2mA applied via 25-35cm^2 (see Nistche et al., 2008). Greater effects are thought to be more likely if stimulation is applied at higher intensities (see Nitsche & Paulus, 2000; Stagg & Nitsche, 2011; Zheng, Alsop, & Schlaug, 2011). For example, 0.6mA must be applied for 5mins to achieve the same after effects seen under 1mA applied for just 3mins (Nitsche and Paulus, 2000); enlarging the reference electrode size to 100cm^2 renders it functionally inert (i.e., 0.01 cm/mA^2; Nitsche et al., 2007); cerebral blood flow increases in the targeted region in line with intensity of anodal tDCS (Zheng et al., 2011); and performance gains are positively related to concentrations of current in targeted left prefrontal regions (Kim et al., 2014). However, there is debate about whether there is a linear relationship between intensity and the size of the neurophysiological/cognitive outcome of tDCS (for this debate, see Fertonani & Miniussi, 2017; Esmaeilpour et al., 2018). Batsikadze, Moliadze, Paulus, Kuo & Nitsche (2013) reported that cathodal tDCS increased excitability when applied at 2mA intensity, and there was no effect for anodal tDCS; Bastani and Jaberzadeh (2013) failed to find a positive correlation between current density and motor cortical excitability increases; whilst Hoy et al. (2013) found that gains on a working memory task were not related to increases in current intensity (see also, Cuypers et al., 2013; Jefferson, Mistry, Singh, Rothwell, & Hamdy, 2009; Kidgell et al., 2013).

1.3.2 Duration

The duration of stimulation influences the duration of after effects and the direction of excitability changes. As I show later, early work on the neurophysiological effects of tDCS showed that by increasing the duration of anodal and cathodal tDCS one can increase the effects post stimulation cessation from a few minutes to roughly one hour to ninety-minutes (Nitsche & Paulus, 2000; for review, see Nitsche et al., 2008). Moreover, in light of this evidence, and as with current intensity, it was assumed that longer durations should boost the efficacy of stimulation. It is the case that for stimulations longer than 4s one can engage plasticity changes that propagate the effects seen during stimulation after stimulation has ceased (Nitsche & Paulus, 2000; Nitsche et al., 2003). However, administering 13 minutes of anodal tDCS followed immediately by another 13 minutes of anodal tDCS (26 minutes in total) can in fact decrease motor cortical excitability (Monte-Silva et al., 2013). These inversions in cortical excitability may reflect a homeostatic response, presumably driven by activation of calcium dependent potassium channels that prevent over-excitation. In support of this view, administering the calcium channel blocker FLZ can return the effect of 26mins anodal tDCS back to excitatory (see Monte-Silva et al., 2013).

1.3.3 Electrode Montage

Since excitability changes are thought to occur under both the anode and the cathode, the site of the reference electrode is important. The *cephalic montage* places the reference electrode above a cortical region that is not implicated in the measured cognitive function or behaviour. This is by far the most popular montage, with the supraorbital area on the side contralateral to the active electrode position being the site of choice across most studies. Another popular cortical site is the homologue of the area underneath the active electrode (sometimes known as the bilateral montage). The intention here is to modify the excitatory-inhibitory balance contralateral cortical regions share (Bloom & Hynd, 2005; Chiarello & Maxfield, 1996; Chrysikou & Hamilton, 2011; Thiel et al., 2006; Silvanto, Muggleton, Lavie & Walsh, 2008; Vines, Cerruti & Schlaug, 2008). However, one major drawback to this montage is that it complicates the interpretability of one's findings, since one is simultaneously modifying the excitability of two cortical sites involved in a given cognitive measure. This problem is avoided all together in a third montage referred to as the extra-cephallic montage. Here, the electrode is placed on areas other than over the brain that are

contralateral to the active electrode, thereby avoiding current flowing to other cortical sites. Typical sites include the cheek (e.g., Ross, McCoy, Wolk, Coslett & Olson, 2010), back (e.g., Ferrucci et al., 2008; Monti et al., 2008), shoulder (e.g., Fertonani, Brambilla, Cotelli & Miniussi, 2014), or neck (e.g., Vandermeeren, Jamart, & Ossemann, 2010), and leg (e.g., Vandermeeren, et al., 2010). However, one concern with this montage is that the electric field strength diminishes as the distance between the two electrodes increases (Miranda, Lomarev & Hallett, 2006; Rush & Driscoll, 1968). A problem the above montages do not address is that the path of current from the anode to the cathode is complex, and diffuses across a wide cortical space due to current shunting via conductive tissue, such as cerebral spinal fluid (CSF) which is found throughout the brain and is highly conductive (Datta et al., 2009; Parazzini, Fiocchi, Rossi, Paglialonga, & Ravazzani, 2011; Salvador, Mekonnen, Ruffini, & Miranda, 2010). This problem is, however, overcome with newly developed 4 x 1 multi-electrode montages (also known as high definition or HD-tDCS). Here a single electrode is surrounded by four smaller electrodes of an opposite polarity (e.g., an anode surrounded by four cathodes), with the surrounding electrodes serving to keep current flow local to the central electrode, thereby delivering a more focalised current. A major drawback to this montage is that you are modulating the activity of one brain region whilst simultaneously modulating in the opposite direction the activity of proximal regions. Given that spatially nearby regions are highly connected relative to distal brain regions (Markov et al., 2014), one might interfere with or cancel out the communication between neighbouring brain regions resulting in a potentially deleterious effect on task performance. This might be particularly the case with a 4 cathode x 1 anode montage, where the net outcome is cortical inhibition given the ratio of cathodes to anodes (Reinhart et al., 2017).

1.3.4 Timing

Stimulation can be applied either before (offline) or whilst (online) cognition or behaviour is measured. Unfortunately, there is as of yet no systematic investigation into the efficacy of stimulation timing, but there is reason to assume that online stimulation is preferable (for similar arguments, see Bikson & Rahman, 2013; Lapenta, Minati, Fregni & Boggio, 2013; Miniussi, Ruzzoli & Walsh, 2010; Miniussi, Harris & Ruzzoli, 2013). Neuroimaging studies report greater cortical excitation during rather than after tDCS administration

(Stagg et al. 2013; Rae, Lee, Ordidge, Alonzo & Loo, 2013), and online effects are also thought to deliver a more focalised effect on cortical tissue, which is important given that between the electrodes the current is diffusely spread (Datta et al., 2009; Parazzini et al., 2011; Salvador et al., 2010). The boost in focality comes from the fact that tDCS causes only minor changes in resting potentials (<1mV) that are below the threshold of excitation, meaning that neurons close to activation are preferentially modulated (e.g., those recruited by a given task; Miniussi et al., 2010). Thus, if applied during a task, tDCS can modulate a specific neuronal population involved in the cognitive task, leading to greater (task-dependent) focality (Bikson & Rahman, 2013; Miniussi et al., 2010; Miniussi et al., 2013) and greater cognitive enhancement (Pisoni et al., 2017).

1.3 Safety of tDCS

The consensus is that risks associated with common tDCS parameters (i.e., 25–35 cm^2 electrodes, currents of 1–2 mA, stimulation durations of 20 mins) are very minimal (for extensive review, see Matsumoto & Ugawa, 2017). Cognitive and behavioural effects are temporary, with common side effects relating mainly to mild skin sensations, such as tingling, itching, pins and needles, and in some cases the feeling of heating or burning. There are extremely rare cases of dizziness, nausea, headache and fatigue (Nitsche et al., 2008); a phosphene (a perceived flash of blue light) when stimulation has been stopped abruptly; or skin sensations akin to a static shock under the electrode.

Neural tissue damage through excito-toxicity, or excessive hyperactivity, does not apply to tDCS because it causes only minor shifts in resting membrane potentials – rather than induce an action potential (see Nitsche et al., 2003; Nitsche et al., 2004). Current density (amperage ÷ electrode size; mA/cm2) and the duration of exposure to this current density, known as total charge (total current density x total stimulation duration, C/cm2), are key safety parameters in this regard. Liebetanz et al. (2009) assessed 10 mins of epicranial electrode stimulation in rats, and estimated that a current density of 1.429 mA/cm2 and total charge of 52,400 C/m2 is sufficient to cause neural damage, although other estimates put the total charge required at 216 C/cm2 (Yuen, Agnew, Bullara, Jacques & McCreery, 1981). These values are well above what is deemed typical in the tDCS field, which are

roughly 0.029 – 0.08 mA/cm2 (current density) and 34–96 C/m2 (total current charge; see Liebetanz et al., 2009; see also, Matsumoto & Ugawa, 2017).

Electrodes present two problems in regards to safety. An electrode is the site of electrochemical toxins, so should not come into direct contact with the skin (Merrill, Bikson & Jefferys, 2005; Nitsche et el., 2008), hence the use of saline soaked sponges or conductive gel, which serve as a buffer at the electrode-skin interface. Another concern is electrode overheating, which can lead to skin damage. The factors that led to cases of skin damage, however, are many, including non-intact skin (e.g., eczema), repeated (daily) sessions of tDCS, and use of tap water (which contains metals that could potentially overheat). Heating can cause mild redness of the skin due to vasodilatation, but this is not linked to skin damage (see Matsumoto & Ugawa, 2017).

Finally, there is precedent for extracephalic montages that place the reference electrode on the leg to cause adverse effects (see Bindman, Lippold & Redfearn, 1964; Redfearn, Lippold & Costain, 1964). However, though it was thought that this was due to the current passing the brain stem, what is more likely is that it resulted from human error (the participant who presented the adverse effects received 10 times the intended amperage). A more recent study found no adverse effects with 20mins of 1mA anodal tDCS applied to the frontal region with the reference electrode placed on the right knee (Vandermeeren et al., 2010), whilst a computer model of the current at the brain stem with this montage to be well within safe limits (0.0045 mA/cm2; see Parazzini, Rossi, Rossi, Priori & Ravazzani, 2013; Parazzini et al., 2014).

1.4 Neurophysiological Effects of tDCS in Healthy Adults

The pioneering work by Michael Nitsche and Walter Paulus (2000, 2001; see also Nitsche et al., 2003; Nitsche et al., 2005) first characterised the neurophysiological effect of tDCS on motor cortical excitability. The authors recorded TMS (transcranial magnetic stimulation) induced MEPs (or motor evoked potentials) before and after tDCS of 1mA applied via 35cm^2 electrodes, with the active electrode placed over the primary motor cortex

and the reference electrode over the supraorbital area[1]. The results showed that compared to pre-tDCS levels just 4s of stimulation led to an increase and decrease of MEPs of up to 20% with anodal and cathodal tDCS, respectively. The MEP amplitude was increased further with greater stimulation durations, with a 40% increase and a 30% decrease in MEP amplitudes respectively under anodal and cathodal tDCS when applied for 5mins, with MEP amplitudes returning to baseline levels 10mins post stimulation cessation. Greater stimulation durations also led to longer after effects, with excitability changes following 13mins of anodal tDCS and 9mins of cathodal tDCS lasting roughly 90 and 60mins post stimulation cessation, respectively (Nitsche & Paulus, 2001; Nitsche et al., 2003).

Later research using various neuroimaging techniques corroborated this early work. Attempts using functional magnetic resonance imaging (fMRI) found that tDCS causes hemodynamic changes (a surrogate for neuronal activity) within targeted regions, with anodal and cathodal tDCS leading to respective increases and decreases in blood flow within targeted primary motor cortical regions (Antal, Polania, Schmidt-Samoa, Dechent & Paulus, 2011) and left prefrontal regions (Holland et al., 2011). These findings are also consistent with positron emission tomography (or PET) scans (see Lang et al., 2005; Paquette, Sidel, Radinska, Soucy, & Thiel, 2011). As mentioned previously, MRS studies showed that levels of glutamate increased whilst levels of GABA decreased under anodal tDCS applied to the primary motor cortex (Clark et al., 2011; Stagg et al., 2009). Studies pairing electroencephalography (EEG) recordings with tDCS show that effects are seen at the network level. Oscillatory activity at lower frequency bands indicative of behavioural inhibition, such theta and delta, are increased with cathodal tDCS, whilst delta is decreased under anodal tDCS (Ardolino, Bossi, Barbieri, & Priori, 2005; Antal et al., 2004; Notturno, Marzetti, Pizzella, Uncini, & Zappasodi, 2014; Pellicciari, Brignani, & Miniussi, 2013; Wirth et al., 2011). Finally, anodal tDCS applied to the primary motor cortex can improve functional connectivity with other motor related areas (e.g., premotor and superior parietal areas), and attenuate connectivity with more distal areas (Polanía, Nitsche & Paulus, 2011;

[1]Motor evoked potentials (or muscular 'twitches'; known as MEPs) are widely considered to be a measure of corticospinal (or motor cortical) excitability, which are usually induced by a single pulse of TMS (transcranial magnetic stimulation). MEPs therefore give an objective measure of the effects of tDCS on cortical excitability.

Polanía, Paulus, Antal & Nitsche, 2011; Polanía, Paulus & Nitsche, 2012), whilst anodal tDCS targeting regions of the frontal lobe improves functional connectivity within the language network (Meinzer et al., 2012; Meinzer et al., 2014a; Meinzer et al., 2014b).

1.5 Cognitive Effects of tDCS in Healthy Participants

The widely held assumption across most studies is that the effects seen at the neurophysiological level translate to the cognitive level. That is to say, an increase in cortical excitation via anodal tDCS is expected to enhance performance whilst a decrease in cortical excitation via cathodal tDCS is expected to worsen performance. However, close inspection of the literature reveals that this is not the rule, especially in studies assessing singles sessions of tDCS in healthy participants. This has been documented in several highly informative meta-analyses and quantitative reviews that attempted to collate published findings in an effort to quantify the general efficacy of tDCS.

Jacobson, Koslowsky & Lavidor (2012) were the first to conduct a meta-analytical review focusing on the anodal-excite/enhance versus cathodal-inhibit/impair (or AeCi) assumption. The authors pooled 34 studies measuring working memory, language or attention and 18 motor studies measuring motor functions and/or neurophysiological effects. The findings indicated 81 versus 47% of cognitive studies reported an AeCi outcome, respectively, which contrasted with 78 versus 87% reported in motor studies. It should be noted that effects from healthy and control participants were pooled together, and because the authors were interested in the AeCi assumption, the chosen inclusion criteria distorted the apparent efficacy of anodal tDCS in cognitive studies, particularly for word production and working memory. For instance, findings that contradicted the AeCi outcome were excluded, non-significant effects were assigned a zero effect size, and in the case of multiple effects reported by one study only the largest effect was considered. This meant that for studies with healthy participants, the ratio of included positive effects versus excluded null effects was 3 versus 26 for word production studies and 4 versus 9 for verbal working memory studies.

Horvath, Forte and Carter (2015a) found similar null effects across roughly 80 studies measuring single sessions of tDCS in healthy participants. Specifically, null effects were

reported across various polarities (anodal or cathodal), cognitive domains (executive functions, language, visual and verbal memory, and miscellaneous higher-cognitive functions), and stimulated areas in the left and right hemisphere (e.g., frontal, temporal, motor and parietal regions). This review has been criticised, however. The authors intended to reduce the impact of idiosyncratic stimulation protocols by excluding outcomes reported by less than two separate laboratories, but this naturally narrowed the number of eligible studies. It also meant that many analyses pooled as few as two or three studies to fit the varying degrees of stringency for each analyses (e.g., by outcome measure, by polarity, by stimulation timing; for further discussion, see Antal, Keeser, Priori, Padberg & Nitsche, 2015; Nitsche, Bikson & Bestmann, 2015; Price & Hamilton, 2015; Price, McAdams, Grossman & Hamilton, 2015). Nonetheless, null effects were reported across all 59 analyses carried out – of which 12 analyses included more than 5 studies.

Dedoncker, Brunoni, Baeken & Vanderhasselt (2016a, but also, see Dedoncker et al., 2016b) reported different results in a meta-analysis of 61 studies coupling singles sessions of left dlPFC stimulation with various cognitive tasks, including language, working memory, executive functions, and visual attention. It was revealed that cathodal tDCS had no effect overall, but anodal tDCS had a generally facilitatory effect across tasks in terms of reaction times and accuracy rates (Cohen's d = -.11 and .18, respectively). However, this last finding must be interpreted with caution for several reasons. First, the authors pooled together effect size estimates from studies testing healthy or patient samples (49 versus 12 studies, respectively). When considered separately, tDCS only improved reaction times but not accuracy rates for healthy participants (reaction times = -.11; accuracy = .04), whilst the opposite outcome with slightly stronger effects were seen for patient samples (reaction times = -.15; accuracy = .22). Second, effect size estimates were based on Cohen's d. This has an upward bias for estimates based on small sample sizes (i.e., $n < 20$; Lakens, 2013), which would inflate the effects reported by a majority of studies included in this meta-analysis (35 out of 49 studies). Third, the authors included in the same meta-analysis multiple effects from multiple dependent conditions (i.e., conditions in which the same participants participated). This violates the independence assumption: the assumption that effect size estimates are drawn from different participant samples, which is assumed by a meta-analysis. Violating this assumption can lead to the underestimation of variance and

potentially inflating the significance of the summary effect size, which is likely given the small sizes of estimated effects (for further discussion, see Chapter 3; for a comprehensive discussion, see Borenstein, Hedges, Higgins, Rothstein, 2009; Lipsey & Wilson, 2001).

Price, McAdams, Grossman, and Hamilton (2015) focused on studies measuring the effects of anodal tDCS on verbal fluency and word learning tasks in an effort to re-evaluate findings by Horvath et al. (2015a). They found that accuracy scores significantly improved with anodal tDCS, either when pooling all studies together (n = 8), studies that applied offline tDCS (n = 4), or verbal fluency studies using offline tDCS (n = 3). However, effects were again small to moderate (roughly 0.5, Hedges' *g*), and may have been inflated by the exceptionally large significant effect size estimates reported by one study measuring word learning (.8, Flöel, Rösser, Michka, Knecht & Breitenstein, 2008) and two studies measuring fluency (1.1, Cattaneo, Pisoni, & Papagno, 2011; .7, Meinzer et al., 2012). For example, in the meta-analysis that pooled three offline fluency studies, two studies reported small non-significant effects (i.e., .2, Cerrutti & Schlaug, 2009; .3, Penolazzi, Pastore & Mondini, 2013a), yet the summary effect was larger and significant (˜.5) – an outcome presumably the result of the larger and significant effect included (1.1, Cattaneo et al., 2011). What is worse, this last effect appears to be difficult to replicate (see Vannorsdall et al., 2016; but for a response, see Cattaneo, Pisoni, Gallucci & Papagno, 2016). Finally, other concerns include that mistakes were made in effect size estimates (Horvath, 2015); the fact that several offline effects reported by one study were excluded; and that studies measuring offline effects immediately post stimulation cessation were pooled with one study that measured the effect 20mins post stimulation cessation (these two last criticisms refer to effects reported by one study: Penolazzi et al., 2013a).

Finally, other meta-analyses that have focused on effects of tDCS on short-term/working-memory tasks show similar inconsistent results. Brunoni and Vanderhasselt (2014) found that reactions times but not accuracy scores improved in eight studies coupling the n-back task with anodal tDCS applied to the dlPFC, but effects in healthy samples (6 out of 8 studies) were significantly smaller than patients (-.14 versus 0.25, Hedges' *g*). Hill, Fitzgerald & Hoy (2016), on the other hand, found that across the ten studies using healthy participants, reaction times and accuracy scores were significantly improved across tasks

including n-back, Sternberg, and span tasks, either when pooling all effect size estimates together (n = 28; Hedges' g = -.15 and .15, respectively) or effects (n = 21) from studies using offline stimulation (Hedges' g = -.16 and .14, respectively), although these were in comparable size to online stimulation where no significant effects were found (Hedges' g = -.12 and .19, respectively). Finally, Mancuso, Ilieva, Hamilton and Farah (2016) conducted a more comprehensive analysis of 31 studies investigating effects in healthy participants, and found no effects of tDCS when pooling studies according to the targeted cortical site (i.e., left/right dlPFC or right parietal lobes), with one exception when pooling studies measuring working memory training whilst targeting the left dlPFC (.29, Hedges' g). In spite of these null findings, Mancuso et al. were also able to detect publication bias, particularly in studies targeting the left/right dlPFC – where there is strong expectation of positive effects. Further, the authors also chose to re-run analyses conducted by Brunoni and Vanderhasselt, (2014), Hill et al. (2015), and Horvath et al. (2015a), including additional studies published since these meta-analyses were published and correcting for effect size estimates which may have inflated significance. The results in every re-analyses showed no significant effect of tDCS.

1.6 Purpose of Thesis

On the face of it, it seems a simple empirical question: can tDCS modify cognition? The answer, however, remains elusive after nearly two decades of research. The several thousand papers published thus far show many positive effects, but outcomes vary and reports of weak and/or null effects are rising. It is imperative that this is reconciled because the cost for not doing so could be great. If the answer to the question is yes, then we hasten the use of tDCS in areas where its application is needed. If the answer is no, then we risk wasting valuable resources and loss of confidence in the scientific rigor of tDCS research from the scientific community, public and grant funders. The question can be answered with a comprehensive re-evaluation of the conditions in which tDCS is and is not reliably effective at modifying cognition. This thesis aimed to understand whether tDCS can modify word production, working memory, and learning, when applied to frontal and temporal regions in healthy participants using conventional stimulation parameters that are thought to produce positive effects (i.e., anodal tDCS applied in a single session). A crucial aspect of the

investigations was a close examination of reasons that explain outcome variation and the gap between the evidence for and against anodal tDCS efficacy. These reasons and the rationale for the tasks I used are detailed in each chapter, but I provide a brief overview below.

1.6.1 Reasons for Inconsistent Effects

1.6.1.1 Protocol

A widely cited reason for the inconsistency in outcomes refers to the choice of stimulation parameters. As previously discussed in Section 1.3, outcomes measuring either cortical excitability or cognition vary depending on the applied current intensity, duration, timing, and electrode montage; a problem compounded by variation in the chosen stimulation parameters across studies (Woods et al., 2016). However, though departures in protocol are important to consider, one can of course always argue that null or inconsistent effects were due to the *wrong* combination of parameters. This is because interpretations of cognitive effects of tDCS are done in the absence of a comprehensive mechanistic framework that can explain or predict outcomes for given conditions, and there is little – if any – reliable evidence indicating conditions in which tDCS is and is not effective (for a comprehensive discussion, see de Berker, Bikson & Bestmann, 2013; Bestmann, de Berker & Bonaiuto, 2015; Fertonani & Miniussi, 2017; Giordano et al., 2017). One crucial aspect of the studies reported in this thesis is the exploration of the parameter space by way of conceptual and direct replications of previous studies, focusing my attention on the most widely used combination of parameters associated with positive outcomes: anodal tDCS applied in a single session in healthy participants.

1.6.1.2 State dependency

Cognitive performance is underpinned by a dynamic balance between inhibition – to maintain stability – versus excitation – to introduce flexibility (or plasticity), which can be conceptualised as an inverted-U shape, with optimal performance at the peak (i.e., balanced excitation-inhibition) and suboptimal performance at the end of each slope (i.e., too much or too little excitation; Cools & D'Esposito, 2011). Baseline differences in the state of cortical excitability, and task-mediated demands on cognitive processes will therefore likely confound the outcome of tDCS (see Monte-Silva et al., 2010; Krause & Cohen Kadosh, 2014; Krause,

Márquez-Ruiz, & Cohen Kadosh, 2013). One possibility is that anodal tDCS may have no effect or even worsen performance if cortical excitability is optimal at baseline, because excitability will be pushed towards too much excitation. An alternative possibility is that anodal tDCS may improve performance if performance is low at baseline, because excitability will be pushed towards the optimal excitation versus inhibition balance. In fact, participants with low baseline performance or cortical excitability levels generally benefit from anodal tDCS across various measures (e.g., mathematics: Sarkar, Dowker & Kadosh, 2014; musical skills: Furuya, Klaus, Nitsche, Paulus & Altenmüller, 2014; picture naming: Ross et al., 2010; working memory: Berryhill, Peterson, Jones & Stephens, 2014; Hsu, Juan & Tseng, 2016; Hsu, Tseng, Liang, Cheng & Juan, 2014; video gaming: Looi et al., 2016; visual perception: Benwell, Learmonth, Miniussi, Harvey & Thut, 2015; Learmonth, Thut, Benwell & Harvey, 2015). The studies reported in this thesis were designed to measure the effect of tDCS on overall performance, but also at the subgroup level to account for individual variation in cortical excitability and its interaction with task mediated load placed on excitation versus inhibition.

1.6.1.3 Power and publication bias

Low power reduces the probability of finding a positive effect, if one truly exists. Yet, it also raises the probability that a significant effect is a false positive with the associated effect being an overestimate of the true effect size – since large effects that occur by variation in the true effect size are more likely to be detected from underpowered studies (for more discussion, see Button et al., 2013a, 2013b; Ioannidis, 2005). Lack of power is now thought to be endemic in tDCS research. The recruitment of 20 participants (or often fewer) is generally regarded to be an acceptable sample size, and according to one reckoning the typical power achieved by tDCS studies is 14%, but can be as low as 4% (Medina & Cason, 2017). False positives can be weeded out overtime, as the accumulated evidence from subsequent studies converges on the true effect size. However, this is only possible when a field is free of publication bias (Button et al., 2013; Schooler, 2001). Publication bias is a problem across a majority of scientific fields, including tDCS (see Mancuso et al., 2016), and results from the reluctance of journals and/or researchers to publish null results, favouring instead novel or positive effects (for review, see Martin & Clarke, 2017), which diverts

attention from replicating published results. Together, lack of power and publication bias can unjustifiably inflate the effectiveness of tDCS. One intention of this thesis was to evaluate the validity of foundational claims made by previous studies by way of replication (conceptual and direct), with the further aim of investigating whether inconsistent outcomes are due to a mixture of chance and publication bias.

1.6.2 Rationale for Tasks

The effects of tDCS were measured on picture naming, word reading, probe (or item recognition) tasks, and novel language acquisition tasks. There are several factors that informed my decision to choose these tasks:

1. They are widely regarded as good indexes of word production and working memory, particularly in regards to executive selection of target representations in the context of competing task-irrelevant information, processes I wish to focus on in order to tap into aspects of state-dependent effects of tDCS (see point 4; see also, Badre & Wagner, 2007; Bialystok, 2009; Jonides & Nee, 2006; Lambon, McClelland, Patterson, Galton & Hodges, 2001; Indefrey & Levelt, 2004; Levelt, Roelofs & Meyer, 1999; Whiteside et al., 2016).

2. They tap cognitive abilities common to cognition in general. These abilities include semantic, phonological/orthographic processing (especially in the case of word production, for reviews, see Jeffries, 2013; Jeffries & Lambon Ralp, 2006; Lambon et al., 2001; Mirman and Brit, 2014), source monitoring, maintenance, activation, and inhibition (especially in the case of working memory, for extensive reviews, see Gathercole & Baddeley, 2014; Johnson et al., 2013; Kiyonaga & Egner, 2014; Thompson-Schill, Bedny & Goldberg, 2005; Novick, Trueswell & Thompson-Schill, 2010; Wilhelm, Hildebrandt & Oberauer, 2013). Thus, these tasks are good proxies to determine the general effectiveness of tDCS across cognition.

3. Several lines of research demonstrate that these tasks are underpinned by focalised regions in the frontal and temporal areas, namely the left inferior frontal gyrus (or LIFG)

and the left posterior temporal gyrus (LpTG, for comprehensive reviews, see Indefrey, 2011; Indefrey & Levelt, 2004; Price, 2000, 2010, 2012). The LIFG is considered to be the location of a domain general executive mechanism that modulates activity of active information so as to bias selection toward task-relevant information whilst suppressing task-irrelevant information, with research suggesting that this function of the LIFG is instrumental for selecting lexical-semantic representation or updating, supressing and selecting contents in working memory (Badre, Poldrack, Paré-Blagoev, Insler & Wagner, 2005; Lau, Phillips, & Poeppel, 2008; Nagel, Schumacher, Goebel, & D′Esposito, 2008; Snyder, Feigenson, & Thompson-Schill, 2007; Thompson-Schill, D′Esposito, & Kan, 1999; Thompson-Schill, D′Esposito, & Aguirre, 1997). The LpTG, on the other hand, is thought to be important for the activation of information from memory, especially semantic memory. In models of word production, the LpTG is considered to be the site of the lexical semantic retrieval (see Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Hickok, Houde, & Rong, 2011) or the neural locus of the *lemma* level in the dominant model of speech production, such as the WEAVER ++ model (see Indefrey & Levelt, 2004; Levelt, 1992; Levelt, Roelofs & Meyer 1999). Thus, the tasks I chose allowed me to concentrate on two specific regions of the cerebral cortex to assess the efficacy of tDCS.

4. The tasks I chose also allowed me to explore potential task mediated interaction with tDCS effects. Models of word production and working memory view these processes as being reliant on the interplay of activation versus inhibitory processes. This means that that anodal tDCS effects might not be uniform across task performance, or always beneficial when targeting the temporal or frontal lobes, but dependent on task related demands on activation versus inhibition. One can hypothesise different outcomes in light of models of word production and working memory, and with reference to the roles played by brain regions described in point 3.

   Dominant models of *word production* view lexical selection as a competitive process in which the target representation is selected in the midst of semantically related alternatives also activated in the course of lexical retrieval (e.g., when naming a picture, the target name and names that are similar in meaning to the target are retrieved at

the same time; see the WEAVER ++ model described by Indefrey & Levelt, 2004; Levelt, 1992; Levelt, Roelofs & Meyer 1999). The most active item is usually selected, but if the activation levels of competitors is high (or similar to the target) then frontally mediated processes external to the lexicon curtails activity of related competitors to ensure they are not selected in error (see Kan and Thompson-Schill, 2004; Moss et al., 2005; Robinson et al., 1998; Wilshire and McCarthy, 2002). Thus, one might hypothesise that in conditions of high competition, anodal tDCS applied to the LpTG may paradoxically worsen lexical selection because it would lead to an increase in activity of – and therefore competition between – active lexical semantic representations. Anodal tDCS to the LIFG, on the other hand, may be selectively effective during cases of high interference from task irrelevant material, where the demand for frontally mediated selection is greatest because there is greater need to inhibit competitors in the service of biasing target selection.

The dominant model of *working memory*, on the other hand, views working memory as multi-component system composed of a temporary store of information that can be actively maintained and manipulated in the service of a given task (see Baddeley, 2002). The temporal constraints on stored information places a premium on frontally mediated executive control processes that select, update and supress contents to inhibit old, irrelevant material to make way for new, task relevant information (Morris & Jones, 1990; (Kessler & Meiran, 2008; Kessler & Oberauer, 2015; Ochsner & Kosslyn, 2013). These processes (particularly inhibition) are tapped by the probe tasks I use, where one must decide whether an item is present in a target list rather than in a previous one. If previous items have not been properly suppressed, it may be falsely recognized as being present in the current list. The task is made more difficult when previously studied items share the same meaning as target information. Thus, though anodal tDCS applied to the LIFG should benefit performance on working memory tasks, by enhancing particularly inhibition, it is likely to be more beneficial during times of effortful inhibition of interference from task-irrelevant material.

As will be described in each empirical chapter (i.e., Chapter 2 and 4), the tasks I chose allowed me to modulate the demand placed on inhibitory and excitatory processes to explore task dependent determinants of the tDCS outcome.

5. There is extensive work examining the effect of tDCS across studies measuring word production, working memory and learning. These are by far the most popular tasks and research areas of choice of researchers across the tDCS literature (for a review of common research areas, see Santarnecchi et al., 2015), and by drawing from extensive literature it gave me plenty of scope to explore and then identify conditions in which tDCS may or may not be effective.

6. tDCS may preferentially operate on learning. It is made clear in Chapter 6 that the effects of tDCS may preferentially operate on representations that are not yet established, as in the case when learning of novel language. Investigating tDCS effects on word learning can therefore be contrasted to the situation in picture naming, word reading, verbal fluency, and probe tasks, where tDCS is operating on a well consolidated network of lexical representations and/or cognitive operations.

1.7 Chapter Summaries

Below are chapter summaries with reference to the main aims and findings of the studies carried out in the course of this PhD research project.

1.7.1 Chapter 2

I carried out four experiments that looked at the effects of tDCS across picture naming and word reading tasks in healthy participants. I used conventional stimulation parameters to target key language areas (i.e., single session anodal tDCS applied to the LIFG or LpTG) during task performance. I measured general picture naming and word reading using tasks that also probed semantic interference effects – a proxy for a change in cortical excitability. Unfortunately, our results failed to show that tDCS modulated overall performance or when considering semantic interference effects. Null effects were also observed even when we controlled for participant variation in response to tDCS effects, and when analysing naming

speeds for difficult items. This investigation was published (see Westwood, Olson, Miall, Nappo & Raffaele, 2017). Following its publication, the paper received a commentary piece by Gauvin, Meinzer and Marcus (2017), which criticized certain aspects of the paper. We wrote a response (Westwood, Olson, Miall & Romani, 2017). Chapter 2 is the submitted manuscript, and I include my commentary response in Appendix 2.5.

1.7.2 Chapter 3

The null findings in Chapter 2 were surprising given the positive effects reported in studies using similar tasks and stimulation parameters. I therefore re-evaluated these findings by carrying out a systematic review and meta-analysis to quantify the effect of tDCS on picture naming and/or word reading speeds and accuracy. Overall, the results showed no evidence of a significant positive effect of stimulation, regardless of whether we pooled studies according to protocol or outcome measure. Linear regression analyses did show that offline stimulation and shorter durations (i.e., less than 15mins) produced significantly greater effect, but effects were still small and difficult to interpret given the number of null effects found. Furthermore, I found evidence suggestive of publication bias that indicated a bias towards reporting positive effects. This review was published (Westwood & Romani, 2017), and forms Chapter 3.

1.7.3 Chapter 4

I moved my focus to assess whether tDCS could modulate performance on verbal fluency and probe tasks. In two studies, I attempted to reproduce positive effects seen on these or similar tasks with anodal tDCS applied to the LIFG with a single application. The results failed to show any significant effect of tDCS both at the group level, the subgroup level (which explored if tDCS effects varied according to task demands and ability), and when analysing naming speeds for difficult items. This chapter was published in *Frontiers of Neuroscience* as part of special issue designed to be a platform for difficult-to-publish null effects.

1.7.4 Chapter 5

The null findings reported in Chapter 4 were in contrast to previous experimental evidence and one meta-analysis that showed positive effects on verbal fluency tasks (see Price et al., 2015). Thus, I carried out a meta-analysis to quantify the effect of single application anodal tDCS on verbal fluency task in studies using healthy participants. This showed a significant outcome, which again was driven by large effects from a minority of underpowered studies. I therefore carried out a replication of one of these studies (i.e., Cattaneo et al., 2011). This replication, however, failed to replicate the result reported by this study. This suggested, therefore, that previous evidence in favour of the efficacy of tDCS in verbal fluency was potentially false positive effects.

1.7.5 Chapter 6

In Chapter 6, I outline a proposal for future research that I intend to undertake, which focuses on two possible reasons for the null effects I reported in the previous chapters. One possible reason is that the effects of tDCS may preferentially operate on representations that are newly formed/ unstable, such as in the case of novel word learning, rather than on representations that have been consolidated through years of practice, such as in the case of word productions and probe tasks. Another reason is that if effects are allowed to accumulate over several repeated applications, cortical excitability and therefore cognition is more likely to be modulated. I report a data that support these potential reasons, but this is only based on a preliminary study.

1.7.6 Chapter 7

Here I summarise the findings of each Chapter, and place them in the wider debate about the efficacy of tDCS when applied to healthy participants in single sessions. I also note potential important limitations of my thesis, and to what extent my findings should (or can) speak to the general efficacy of tDCS in conditions not investigated in my research. I also offer recommendations to improve the scientific rigour of the field, placing emphasis on the need to replicate and confirm previous results as part of broader aim to make tDCS research more transparent.

Chapter 2: Investigating the effects of tDCS on picture naming and word reading

2.1 Overview

In this chapter, the goal was to investigate whether a single session of anodal tDCS applied to the left inferior frontal gyrus and left middle posterior temporal regions can modulate lexical selection during picture naming and word reading tasks in healthy participants. As previously discussed in the introductory chapter (see Section 1.6.1 Reasons for Inconsistent Effects), variation in tDCS outcomes may be attributable to the confounding impact of concurrent cortical activity (see Miniussi et al., 2013). Models of language production state that naming involves cortical excitation to activate the target representation and suppression of activity to curtail excitation of related words that may be selected in error. Thus, instead of only looking at an overall effect of tDCS, I chose to assess whether the increased excitability induced by anodal tDCS is overall positive versus negative depending on the lexical mechanisms (activation vs selection) primarily required by the task. I therefore carried out analyses to assess the overall effectiveness of tDCS, but also additional analyses that looked at performance when considering the outcome direction, at differing speeds, and semantic interference effects. I submit below the final published version[2]. Following its publication, the paper received a commentary piece by Gauvin, Meinzer & Marcus (2017)[3], so I submit the final published version[4] of my response in Appendix 2.5.

2.2 Introduction

Transcranial direct current stimulation (tDCS) is a popular technique for modifying cognition using a weak electric current. Over the past decade, thousands of articles have

---

[2]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.

[3]Gauvin, H. S., Meinzer, M., & de Zubicaray, G. I. (2017). tDCS effects on word production: Limited by design? Comment on Westwood et al. (2017). Cortex.

[4]Westwood, S. J., Olson, A., Miall, R. C., & Romani, C. (2017). tDCS modulation of naming in healthy participants: Negative results and still no explanation–a response to a commentary by Gauvin et al.(2017). Cortex.

reported beneficial effects especially in language tasks in participants with healthy (Prehn & Flöel, 2015) and pathological brains (for aphasia, see de Aguiar, Paolazzi, & Miceli, 2015; for dyslexia see, Heth & Lavidor, 2015). Based on early research on the motor cortex, cortical excitability can be modulated via shifts in resting membrane potentials, resulting in hypopolarization/excitation versus hyperpolarization/inhibition depending on the polarity of stimulation (i.e., anodal versus cathodal). However, cognitive effects are far more complex and unpredictable (Horvath et al., 2015a). This is in part because tDCS effects interact with ongoing cortical activity (see Silvanto et al., 2008), as indicated by the general effectiveness of tDCS in patient samples (for review, see de Aguiar et al., 2015; Cappon, Jahanshahi, Bisiacchi, 2016). It may therefore be that tDCS can modulate cognition in pathological brains where excitability or processing capacity is unusually low or dysfunctional, but not in healthy brains where neuronal excitability is operating at optimal levels. If true, this will limit the applicability of tDCS. We aimed to gather further evidence on this question by focusing the effects of single-session, anodal tDCS in normal participants coupled with picture naming and reading tasks, and by considering the moderating influence of cortical excitability resulting from individual differences and task demands.

The reliability of tDCS in cognitive tasks has been questioned in recent reviews. Horvath et al. (2015a) found no evidence of any cognitive effects across eighty studies on healthy participants using single sessions of tDCS. In a companion review, Horvath, Forte, and Carter (2015b) also showed no *neurophysiological effects* of tDCS beyond the modulation of motor evoked potential (MEP) amplitudes. Meta-analyses focusing on working memory/short-term memory effects in healthy samples reported similarly significant but small effects of anodal tDCS (e.g., Brunoni & Vanderhasselt, 2014; Hill et al., 2015). For example, Dedoncker et al. (2016a) found a significant but unimpressive reduction in response times following single sessions of anodal (or excitatory) tDCS applied to the left dorsolateral prefrontal cortex in healthy volunteers (effect size: -.11). However, a recent and arguably more comprehensive review by Mancuso et al. (2016) focusing on the effects of anodal tDCS in healthy participants revealed that effects became non-significant after correction for publication bias. This is important given the notorious "file-drawer" tendency to favour publishing studies reporting significant results.

Only one published review has examined effects of tDCS on language tasks in healthy participants, and it has not included naming tasks. Price et al. (2015) examined effects in verbal fluency (n = 6) and word learning (n = 2) and found a small anodal tDCS improvement in accuracy scores when all studies were pooled together, but also when analyses were limited to the four studies using offline stimulation (i.e., applied prior to task performance) or the three studies measuring offline effects in verbal fluency. Here as well, however, effects were small (less than roughly .05), and depended largely on two studies with abnormally large effects (˜.8; Flöel et al., 2008; ˜1.2; Cattaneo et al., 2011). What is worse, the effect in one of these studies (i.e., Cattaneo et al., 2011) has not been replicated since (see Penolazzi et al., 2013a; Vannorsdall et al., 2016; but see Cattaneo et al., 2016 for response). Another review by Jacobson et al. (2012) showed no cathodal-induced decrements for language studies (0 out of 5 studies), but significant anodal induced improvements (7 out of 8 studies). This review, however, included both patient and control samples. Moreover, since the aim was comparing cathodal and anodal stimulation, for each study, only the most significant effect for either cathodal or anodal stimulation was included across conditions, a zero effect size was assigned to null outcomes, and any effect that contradicted an anodal-excitation/ cathodal-inhibition outcome was excluded. In actuality, across the four studies investigating language production in healthy participants, only 3 out of 26 effects were significant.

Variation in tDCS outcomes may be due to methodological differences across studies, especially in terms of the parameters of the applied current (for further discussion, see Antal et al., 2015; Horvath, Carter, & Forte, 2016; Nitsche et al., 2015), but also to interaction with ongoing cortical activity (see Miniussi et al., 2013). Picture naming could be an important task to assess these interactions. Naming involves both the need for cortical excitation to allow retrieval of target representations and the need to curtail excitation of related words that may otherwise reach 'activation threshold' and be produced in error (for similar argument, see Miniussi et al., 2013). Depending on the task, one can have a relatively greater need of activation/excitation versus selection/control. Therefore, instead of looking at an overall effect of tDCS, one can assess whether the increased excitability offered by tDCS is overall positive versus negative depending on the lexical mechanisms (activation *vs*

selection) primarily required by the task. A crucial feature of our investigation will be to look at these potential differences.

The interplay of lexical activation and selection in word retrieval is well demonstrated with paradigms where the presence of semantically related words increases the need for mechanisms of selection and results in longer time/less accuracy in retrieving the target word. This so-called semantic interference effect is demonstrated when: a) naming pictures in the presence of semantically related versus unrelated words (picture-word interference; Abdel Rahman & Melinger, 2007; Belke & Stielow, 2013; Levelt, Roelofs, & Meyer, 1999; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007), b) repeatedly naming sets of semantically related versus unrelated words (cyclic blocked picture naming; Belke, 2013; Belke & Stielow, 2013; Oppenheim, Dell, & Schwartz, 2010; Schnur, Schwartz, Brecher, & Hodgson, 2006), c) comparing naming of exemplars early in a sequence of related pictures – when interference is low – with naming exemplars later in the sequence – when interference has built up (continuous naming paradigm; Belke, 2013; Belke & Stielow, 2013; Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Effects in picture naming are sometimes compared with effects in reading with the expectation that difficulties with lexical-semantic selection will affect picture naming, but not reading, where targets are retrieved from an orthographic rather than a semantic specification (see Belke, 2008, 2013).

One can put forward different hypotheses on how tDCS could modulate effects of semantic interference. One may assume that *anodal* tDCS, which increases excitability, will improve performance when retrieving words in neutral conditions, but will have more mixed effects when retrieving words in the face of competitors. In this context, effects can even be negative, because it is harder to select among highly activated competitors (i.e., interference effects will *increase*). Furthermore, these contrasting effects may depend on the site of stimulation. It has been suggested that negative effects of anodal tDCS are more likely when applied to temporal areas, which are involved in lexical activation and retrieval (e.g., Indefrey & Levelt, 2004; Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014), while positive effects may be more likely when anodal tDCS is applied to the frontal lobe, which are involved in boosting mechanisms of control and selection (e.g., Hirshorn & Thompson-Schill, 2006; Novick et al., 2010; Scott & Wilshire, 2011). Note, however, that this further

hypothesis depends on two controversial assumptions: 1. that effects of tDCS can be focal enough to target specifically one of two adjacent cortical areas (but see Datta et al., 2009); 2. that top-down frontal mechanisms contribute to lexical selection in addition to mechanism of lateral inhibition intrinsic to the lexical module (see Hamilton & Martin, 2005, 2007 for a discussion).

Pisoni, Papagno, and Cattaneo (2012) tested effects of tDCS on semantic interference using a cyclic blocked picture-naming paradigm. As predicted, they found increased interference following stimulation of the temporal lobes, but decreased interference following anodal tDCS of the frontal lobe. Meinzer, Yetim, McMahon, and de Zubicaray (2016) and Wirth et al. (2011) also found decreased interference during frontal tDCS with the same paradigm. However, Meinzer et al. (2016) did not replicate the expected increased interference following temporal stimulation and Henseler, Mädebach, Kotz, and Jescheniak (2014) found no significant effect of either frontal or temporal stimulation with a picture-word interference paradigm. These findings, together with more general reviewed findings, point to the limited efficacy of single session tDCS to modulate cognition in healthy participants. In our experimental study, we want to try to replicate these findings, but also explore reasons for variability by considering how tDCS effects may interact with individual differences in cortical excitability.

Participants are likely to differ in baseline levels of cortical excitability for a variety of factors (for extensive reviews, see Krause et al., 2013; Li et al., 2015). If cognitive performance depends on an optimum level, with worse performance associated with either too low or too high excitability, then some individuals may show improvement after anodal tDCS, whilst others may show no effect or even worse performance depending on baseline levels. Individual variability in response to both TMS (Silvanto et al., 2008) and tDCS (López-Alonso, Cheeran, Río-Rodríguez & Fernández-del-Olmo, 2014; Wiethoff, Hamada & Rothwell, 2014) has been demonstrated in the motor domain. López-Alonso et al. (2014), for example, reported that following tDCS more than half of participants showed no increase in TMS-elicited MEPs, but actually a slight decrease. There are also indications that tDCS effects may depend on baseline level of performance (Hsu et al., 2014; Tseng et al., 2012). For example, Tseng et al. (2012) showed that anodal tDCS induced improvements in visual

short-term memory and associated increases in event-related potentials (ERPs), but that both of these changes were limited to participants with initially poor performance. These individual sources of variability may compound task-mediated variability in producing variable tDCS outcomes.

In our experimental investigation, we will use naming and reading tasks to assess effects of tDCS both overall and, more specifically, on interference effects. We will use 'best practice' anodal stimulation protocols. With cyclic blocked naming picture, we will target frontal areas; with continuous naming, we will contrast stimulation of frontal and temporal areas. Frontal stimulation may be particularly helpful to reduce interference effects, boosting selection mechanisms, which control the activation of potential competitors. Temporal stimulation, instead, may increase the activation of competing items, leading to even stronger interference.

In addition, we will consider the possibility of individual variation. Individuals with high baseline levels of excitability may be more likely to exceed an optimal level of activation, especially in naming conditions where a sequence of competitors increases overall activation levels. To evaluate potential effects of tDCS which may have a different sign (positive or negative) in different individuals, we will consider absolute (independent of sign) inter-session differences in an experimental group, where one session is carried out with real stimulation and one with sham stimulation. We will, then, compare these differences with absolute inter-session differences in a control group, where both sessions are carried out in neutral, no stimulation conditions. If tDCS has any effect, differences in the experimental group, due to tDCS, should be larger than differences in the control group, due to random variability between sessions.

Finally, we will also look at effects of tDCS depending on item variability. We will carry out so-called Vincentized analyses where the RTs of each participant are separated into different bins according to the irrelative speed (very slow, slow, fast, very fast; for a similar method, see Henseler et al., 2014) and then assess the effects of tDCS for each bin. RTs in the 'very slow' category may be particularly susceptible to modulation by tDCS (see also Ross et al., 2010).

2.3 Method

2.3.1 Experiment 1: continuous picture naming and reading

Experiment 1 assessed effects of tDCS on picture naming by applying anodal tDCS to frontal (Experiment 1A and 1B) or temporal areas (Experiment 1C). Following Pisoni et al.'s (2012) logic, we expected frontal anodal tDCS to facilitate naming by boosting the ability to select the target word amongst competitors, but temporal stimulation to have possible negative consequences by increasing competition among related items. Differently from Pisoni et al. (2012), however, we used a continuous naming task where participants are presented with sequences of semantically related pictures, but are generally not aware of relationships between pictures because items belonging to the same semantic category are intermixed with distractors. This makes the disruptive effect of competitors less susceptible to strategic control. A reliable increase of RTs for every new item belonging to the same category in a sequence has been shown across studies (with increases of as much as 30 msec for every additional picture; e.g., Belke, 2013; Belke & Stielow, 2013; Howard et al., 2006).

We paired picture naming tasks with corresponding reading tasks to see whether interference effects were specific to the semantic domain and to test more general facilitation effects in word production. If tDCS selectively modulates interference effects in picture naming, with no interference effects in reading, this will show that there are specific effects of tDCS on lexical-semantic control.

2.3.1 Experiment 1A

2.3.1.1 Tasks

Participants carried out *word reading* and *picture naming* tasks, with picture names corresponding to the words used in reading. Stimuli were presented one by one on a computer screen, and participants named stimuli as fast and as accurately as possible. In both tasks, the experimental pictures/words belonged to sets of semantically related items, with related items being separated by a variable number of unrelated items. We measured general speed and accuracy of performance, but also accumulation of semantic interference effects across sets of related pictures.

2.3.1.2 Design

Each participant carried out both tasks in each of two testing sessions, scheduled one week apart and involving parallel versions of the same tasks. In the experimental group, sham stimulation was applied in one session and real stimulation in the other. In the control group, no stimulation was applied in either session. Reading was always done first in order to prime and, therefore, facilitate retrieval of picture names. The order of real and sham stimulation sessions, and which particular version of the task was paired with each session, was counterbalanced across participants. Reading lasted for 5-6 min and picture naming for 9-10 min. Stimulation covered all testing times. It started at the beginning of the reading task, and was applied continuously with no gap when the task was changed.

2.3.1.3 Stimuli

165 coloured pictures (720 x 540 pixel dimensions) were taken from a variety of sources, and the same number of corresponding words made up the stimuli. 120 stimuli were experimental and 45 were "fillers". Experimental stimuli were drawn from 24 semantic categories, with 5 members to each category (for a listing see Appendix 2.1). Presentation of stimuli followed Howard et al. (2006): the first and last five items were filler items; pictures from the same category were presented in a sequence that separated category members by 2, 4, 6, or 8 items composed of fillers or pictures from other categories; each of the 24 categories used a different sequence of lags. The parallel versions of the tasks included the same categories, but different items. To make sure that positional effects were not confounded with other variables, items in different positions were carefully matched for typical age of acquisition (Kuperman, StadthagenGonzalez, & Brysbaert, 2012), frequency (based on CELEX Database; Baayen, Piepenbrock, & Gulikers, 1995), word length, and name agreement[5]. These variables were also matched across the two versions of the task (Appendix 2.2).

2.3.1.4 Task Procedure

[5]Fifteen undergraduate students were shown the 165 pictures and were asked to name each picture. The experiment was self paced. Name agreement was measured in terms of the number of different names given to each picture. For example, low name agreement would mean relatively more alternatives, and visa versa.

Participants were verbally instructed to read or name the stimuli as fast and as accurately as possible, and to use sub-ordinate nouns (e.g., correct responses to water-lily could be "water-lily" or "lily" but not "flower"). A practice task familiarized participants with the voice key. Each naming/reading trial started with the presentation of a fixation cross for 1000 msec followed by a blank screen for 250 msec. Stimuli were then presented centred, for 2500 msec or until the participant made a response. A blank screen followed for 500 msec before the next trial started. Stimuli were presented using E-Prime 2 Software and a Dell Laptop computer screen (screen size: 15.6"). Words were presented in Arial typeface 24-font. Vocal responses were recorded using a Sony ICDPX333.CE7 voice recorder. The voice key was a serial response box (Refresher Detector System, Psychology Software Tools, INC). The microphone was a Sony ECM-MS957.

2.3.1.5 tDCS

tDCS was administered using a battery driven NeuroConn DC-Stimulation via a pair of saline soaked sponges. Stimulation was administered using a double-blind procedure, whereby both the experimenter and the participant were unaware of the type of stimulation administered in a given session. For sham stimulation, an intermittent current of 110 mA was delivered for a period of 3 msec every 550 msec. This produces the perceptual sensations of real stimulation without modulating underlying brain areas (Palm et al., 2013). For real stimulation, a constant current of 1 mA was administered for 15 mins with a ramp up and ramp down of 30 sec to reduce discomfort and perceptual differences with sham stimulation. The active electrode (9 cm^2; current density ¼ .11 mA/cm^2) was placed over the *left inferior frontal gyrus* (LIFG) whilst the reference electrode (35 cm^2) was placed over the contralateral supraorbital area. The LIFG was located by measuring 2 cm from the corner of the eye towards the preauricular point of the left ear then 3 cm upwards perpendicular from this measurement, which corresponds to F7 using the electroencephalogram (EEG) 10/20 position system (Devlin & Watkins, 2007). At the end of each session, participants completed a feedback questionnaire (see Fertonani, Rosini, Cotelli, Maria, & Miniussi, 2010) to assess the effectiveness of stimulation blinding.

2.3.1.6 Participants

Fifty undergraduate students from Aston University participated for course credits or financial reimbursement, and were assigned to the experimental or control group in a semi-random fashion. Two participants in the experimental group and control group failed to attend the second session due to other commitments. This left eighteen participants (10 female; 21 ± 2.76) in the experimental group and twenty-eight participants (17 female; 23 ± 2.52) in the control group. All participants were right-handed and native English speakers. We excluded volunteers with language impairments, history of migraine, headaches (frequent or severe), skin disorders (e.g., eczema), any adverse experience to previous tDCS, any history of epilepsy or stroke, head/metal implants, any neurological disorders, and any volunteers who had participated in a tDCS or TMS study in the 6 months prior to the current study.

2.3.2 Experiment 1B

As shown later, Experiment 1A returned no evidence of tDCS effects. Therefore, we changed the stimulation protocol to increase the chances of positive effects as detailed below. In all other methodological aspects, Experiment 1B was the same as Experiment 1A.

2.3.2.1 Stimuli

In Experiment 1A, the order of stimuli was the same for each participant. In Experiment 1B, we created 24 different stimuli orders for each of the two matched versions of the naming (and reading) task, with a different sequence of lags for the different semantic categories, but most importantly with a different set of items in the five positions. Each participant was administered one of these 24 versions (for a similar procedure, see Howard et al., 2006). This was to ensure better counterbalancing of items across positions.

2.3.2.2 Procedure

The order of reading and naming tasks was counterbalanced across participants instead of reading always coming first.

2.3.2.3 tDCS

We increased the intensity of the current from 1 mA to 1.5 mA, and increased the size of the active electrode from 9 to 25 cm^2. These changes were made to reduce current density

(e.g., .06 mA/cm^2 instead of .11 mA/cm^2); larger electrodes may make the current more uniform and increase cortical excitation (Miranda et al., 2006). Stimulation duration was increased by 10 mins (total stimulation duration now 25 mins), with a 5 min delay added between the onset of stimulation and the experimental tasks (during which participants read the instructions again from the computer screen) to ensure tDCS effects were fully engaged at task initiation (see Nitsche & Paulus, 2000; Nitsche et al., 2008; Price et al., 2015).We also added 5mins at the end to ensure that both tasks were covered by stimulation. Two participants in Experiment 1A had completed naming slightly after stimulation offset (these participants were, in any case, excluded from analysis because they failed to show up to the second session).

2.3.2.4 Participants

Thirty-nine undergraduate students from Aston University participated for course credits or financial reimbursement. Data from four participants in the experimental group were lost due to a technical problem. Thus, the final experimental group included twenty participants (12 female; 21±2.92) and the control group twenty-five participants (13 female; 21±3.73).

2.3.3 Experiment 1C

In Experiment 1C, we assessed whether contrasting effects of tDCS would be found with temporal lobe stimulation. In all methodological details, bar those reported below, Experiment 1C was the same as Experiment 1B.

2.3.3.1 tDCS

The active electrode (25 cm^2) was placed over the *left mid-posterior temporal lobe* area (pMTG) whilst the reference (35 cm^2) was placed over the contralateral cheek. The pMTG was determined to be at the halfway point between T3 and T5 using the 10e20 International EEG system. We used the contralateral cheek for the reference electrode as it was speculated that by doing so we can avoid current flow through frontal areas, thereby avoiding the difficulty in localizing possible behavioural effects.

2.3.3.2 Participants

Eighteen (13 female; 19.8 ± 2.8) from Aston University participated for course credit or for financial reimbursement. No participants were allocated to the control group as control data from Experiment 1B also applied to 1C.

2.4 Experiment 2: cyclic blocked picture naming

In Experiment 2, tested the effects of tDCS on cyclic blocked picture naming. This paradigm has been extensively studied (for a review, see Belke & Stielow, 2013), and positive effects of tDCS have been reported (Meinzer et al., 2016; Pisoni et al., 2012; Wirth et al., 2011). In this paradigm, participants are asked to repeatedly name sets of pictures that are either semantically related or unrelated. There is, initially, a marked facilitation, with reaction times falling in cycle 2 relative to cycle 1, due to practice. The facilitation continues in subsequent cycles, but the magnitude of this facilitation is reduced for sets of semantically related pictures, due to increased interference amongst competitors which counters facilitation effects. Even more than the previous continuous naming task, this task taps into the ability to select between a set of highly activated lexical representations, because the same small set of pictures is presented repeatedly over a number of cycles. Consistent with this view, imaging evidence shows increased prefrontal activity, presumably linked to the effort for selection, during cyclic blocked picture naming (Schnur et al., 2006), and improvement during anodal tDCS stimulation is associated with increased activity in frontal areas (Wirth et al., 2011).

2.4.1 Task

Participants named as fast and accurately as possible sets of six pictures, with pictures presented one at a time and each set presented four times in a row (four cycles). We measured general naming speed and accuracy, and semantic interference as it builds up across repeated cycles.

2.4.2 Design

Participants carried out two testing sessions in different stimulation conditions (real or sham), one week apart, with parallel sets of materials. The order of real and sham

stimulation, and the task version coupled with each type of stimulation, were counterbalanced across participants. The task lasted for roughly 20 min. Stimulation began five minutes before participants initiated the task and lasted the entirety of the task. During the 5 min delay, participants read task instructions via a computer screen.

2.4.3 Stimuli

72 black and white line drawings were taken from the Snodgrass and Vanderwart (1980) set. Pictures were grouped into 12 sets of six pictures: half the sets included *semantically related* pictures, the other half included semantically unrelated pictures created by selecting one member from related sets (see Appendix 2.3). Pictures were presented in 4 cycles in different quasi-random orders (i.e., each picture occupied a different ordinal position across the 4 cycles, and the last item of a cycle and the first of the following cycle were never the same). The related/unrelated blocks were also alternated in a quasi-random order to ensure that no more than two blocks of the same type were shown consecutively. The order of stimulus presentation was the same for all participants. The two versions of the tasks included different semantic categories and different items. Items in the two versions were carefully matched for age of acquisition (Kuperman et al., 2012), frequency (based on CELEX Database; Baayen et al., 1995), word length and name agreement (based on H statistic from Snodgrass & Vanderwart, 1980; see Appendix 2.4).

2.4.4 Procedure

Participants were given the same instructions as in Experiment 1. Additionally, they were familiarized with the pictures before beginning the experiment. They were first presented with each picture with its name written below, and then with the pictures on their own and asked to name them. An accuracy score of 90% or more was needed to progress to the main experiment.

In the main experiment, each naming block began with a "Get Ready ..." message for 4000 msec, followed a blank screen for 1000 msec and then a fixation cross for 1000 msec. The picture was then presented and remained on the screen until the participant gave his or her naming response. The end of each block of pictures was followed by blank screen for 1000

msec, and by an "End of block ..." message which requested the participant to "Press any button" to start the next block. Stimuli were presented using E-Prime 2 Software. Vocal responses were recorded using a TASCAM DR-680 digital voice recorder with a Rode NTG 2 Condenser Shotgun Microphone. Vocal response times were measured using a Cedrus SV-1 voice key.

2.4.5 tDCS

The stimulation protocol matched Experiment 1B in every way except that stimulation was administered using a battery driven Eldith DC-Stimulation device (functionally equivalent to the Neuroconn DC stimulator).

2.4.6 Participants

Thirty-two undergraduate students from University of Birmingham participated for course credits or for financial reimbursement. A technical error meant that data from three participants in the experimental group had to be excluded, leaving seventeen participants (12 female; 21 ± 2.40) in the experimental group and thirteen participants (7 female; 22 ± 1.76) in the control group.

2.5 Across experiments

2.5.1 Ethical approval

Our experimental investigation was approved by The Ministry of Defence Research Ethics Committee, by the Aston Research Ethics Committee and by the University of Birmingham Ethics Committee. All participants gave written informed consent prior to any testing session.

2.5.2 Scoring

Response accuracy was scored after each testing session. Only near-synonyms (e.g., "Hoover" instead of "vacuum") were allowed as correct, any other response was scored as incorrect. Incorrect responses were excluded from RT analysis, as well as RTs below 250 msec and above 2.5 standard deviations from the participant mean. For picture naming, we analysed

percentage error rates and RTs. Errors rates were not analysed for word reading and cyclic blocked naming tasks because they were very low (<5% and <7%, respectively).

2.5.3. Data re-sampling

In the experimental groups, the order of stimulation (i.e., Sham *vs* Real) and the set of stimuli (i.e., A *vs* B) were counterbalanced. So, in the first session, half of the participants received sham whilst the other half received real stimulation, and half of the participants that received either type of stimulation saw stimuli set A whilst the other half saw set B. In the control group – where stimulation was not applied – half of participants saw set A in the first session and B in the second, and vice versa. To make results from the control group comparable with results from the experimental group, we resampled control data to create two pseudo datasets for sessions 1 and 2, so-called pseudo-sham and pseudo-real so that the order of presentations (session 1 vs 2) and stimulus set (A vs B) was also counter-balanced across these two sessions.

2.5.4 Data analysis

Data was analysed with repeated factor ANOVAs (analysis of variance) to assess the effect of condition in the experimental (*tDCS vs Sham*) and control (*Pseudo-tDCS vs Pseudo-Sham*) groups separately. In addition we ran mixed factor ANOVAs, which combined data from both groups, and considered group as a between-participants factor. This provided a more rigorous test. If tDCS were to have an effect, we excepted an interaction between condition and participant group because the experimental group would show a significantly larger effect of condition than the control group – where stimulation was not applied. For these analyses, we report only the condition by group interactions, since the main effect of condition is irrelevant.

2.6 Results

2.6.1 tDCS feedback questionnaire

Participants tolerated stimulation well. None reported adverse effects nor withdrew from the study because of stimulation. Common sensations were pinching, itching, burning and heat, all with mild to moderate intensity. These sensations differed significantly between

stimulation conditions for some participants, but not systematically across experiments or conditions. When asked to identify what form of stimulation they received, participants reported to be guessing or using a 'gut feeling'. Repeated samples t-tests showed that correct guesses never exceeded chance level (Exp 1A: $F(1,17) = .32$, $p = .58$, $\eta_p^2 = .02$; Exp 1B: $F(1,19) = .32$, $p = .58$, $\eta_p^2 = .02$; Exp 1C: $F(1,17) = .14$, $p = .72$, $\eta_p^2 = .01$; Exp 2: $F(1,16) = 1.00$, $p = .33$, $\eta_p^2 = .02$).

2.6.2 Overall effects of tDCS

Effects of stimulation across tasks, experiments and participant groups are shown in Figure 2.1. We carried out individual one-way ANOVAs for each experiment and participant group to assess whether there was an effect of *Condition* (*tDCS vs Sham* for experimental group; *Pseudo-tDCS vs Pseudo-Sham* for control group). In no experiment was there a significance effect of *Condition*, either in picture naming RTs (across groups: $F < 1.4$, $p > .25$, $\eta_p^2 < .08$), errors ($F < 1.33$, $p > .26$, $\eta_p^2 < .05$), or reading RTs ($F < 1.05$, $p > .32$, $\eta_p^2 = .06$), see Fig. 1.

Mixed factor ANOVAs combined results across experiments and participant groups, with *Group* (*Experimental vs Control*) and *Task* (*Continuous Naming vs Cyclic Blocked Naming*) as between-participants factors and *Condition* (*tDCS vs Sham* for experimental group; *Pseudo-tDCS vs Pseudo-Sham* for control group) as a within-participants factor. For picture naming RTs, there was no main effect of *Group* ($F(1,135) = .002$, $p = .97$, $\eta_p^2 = .00$), but a significant main effect of *Task* ($F(1,135) = 154.55$, $p < .001$, $\eta_p^2 = .53$), with faster RTs in cyclic blocked naming, as expected. There were no significant interactions, including *Group* x *Task* ($F(1,135) = .05$, $p = .83$, $\eta_p^2 = .00$), *Condition* x *Task* ($F(1,135) = 1.1$, $p = .30$, $\eta_p^2 = .01$) and, crucially, *Condition* x *Group* ($F(1,135) = .12$, $p = .73$, $\eta_p^2 = .00$) or *Condition* x *Group* x *Task* ($F(1,135) = .01$, $p = .93$, $\eta_p^2 = .00$). For picture naming errors, *Task* was not a factor because there were not enough errors to analyse in cyclic blocked naming. There was a main effect of *Group* ($F(1,107) = 8.46$, $p = .004$, $\eta_p^2 = .07$), with the control group being more error prone than the experimental group ($M \pm SE$: 16 ± 1% vs 13 ± 1%). There was, crucially, no *Condition* x *Group* interaction ($F(1,107) = 1.76$, $p = .19$, $\eta_p^2 = .02$). For reading RTs, there was a main effect of *Group* ($F(1,107) = 4.00$, $p = .05$,

η_p^2 = .04) with the experimental group being slower than the control group (524 ± 8 vs 500 ± 9), but no *Condition* x *Group* interaction ($F(1,107)$ = .52, p = .47, ηp2 = .01).

These results show no systematic effects of tDCS. There were some significant differences between the experimental and control group. The experimental group was faster in naming, but slower in reading than the control groups. It is possible that stimulation (both real and sham) modulates level of performance, but more detailed interpretations are difficult.

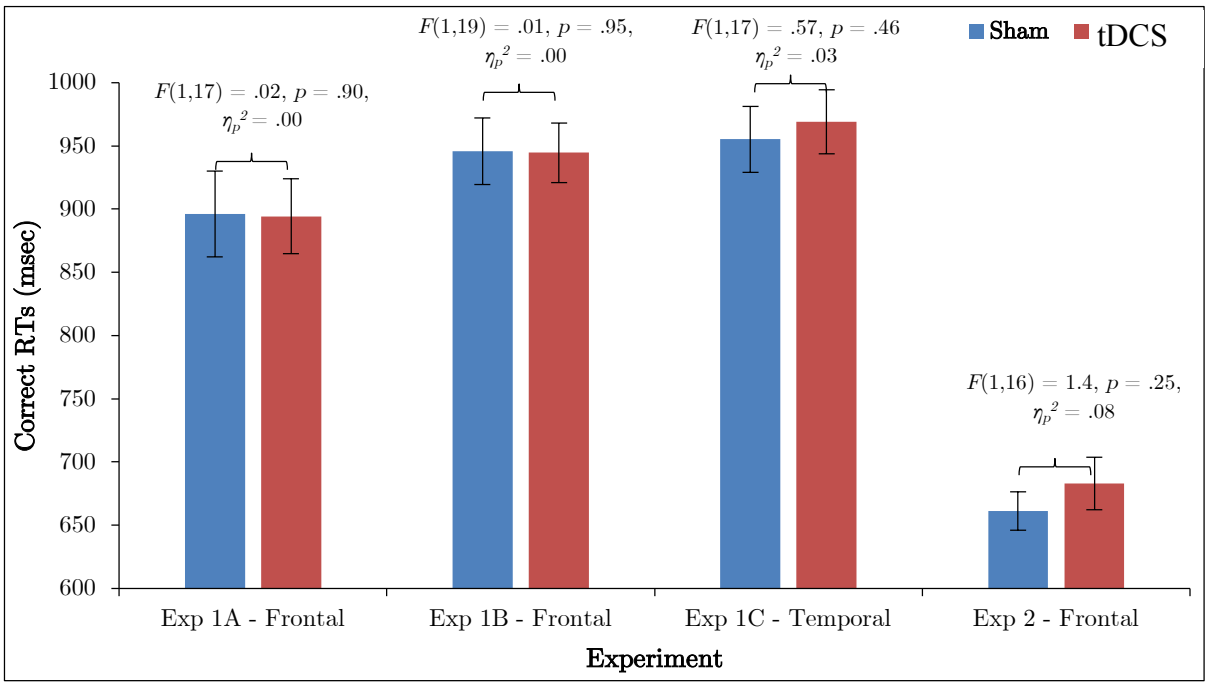
2.6.2.1 Interaction with cortical loci of stimulation

To test for a possible interaction between stimulation site and tDCS, for the experimental group only we conducted a mixed factor ANOVA, with *Site* (*Temporal vs Frontal*) as a between participants factor and *Condition* (*tDCS vs Sham*) as a within participants factor. We report, here, only experiments 1B and 1C, which used exactly the same paradigm. There were no main effects of *Site* (naming: $F(1,36)$ = .25, p = .62, η_p^2 = .01; errors: $F(1,36)$ = 1.71, p = .20, η_p^2 = .05; reading: $F(1,36)$ = .001, p = .97, η_p^2 = .00) and *Condition* (naming: $F(1,36)$ = .26, p = .62, η_p^2 = .01; errors: $F(1,36)$ = .07, p = .79, η_p^2 = .00; reading: $F(1,36)$ = .01, p = .92, η_p^2 = .00), nor a *Site* x *Condition* interaction (naming: $F(1,36)$ = .36, p = .55, η_p^2 = .01; errors: $F(1,36)$ = 1.01, p = .32, η_p^2 = .03; reading: $F(1,36)$ = .10, p = .75, η_p^2 = .00).

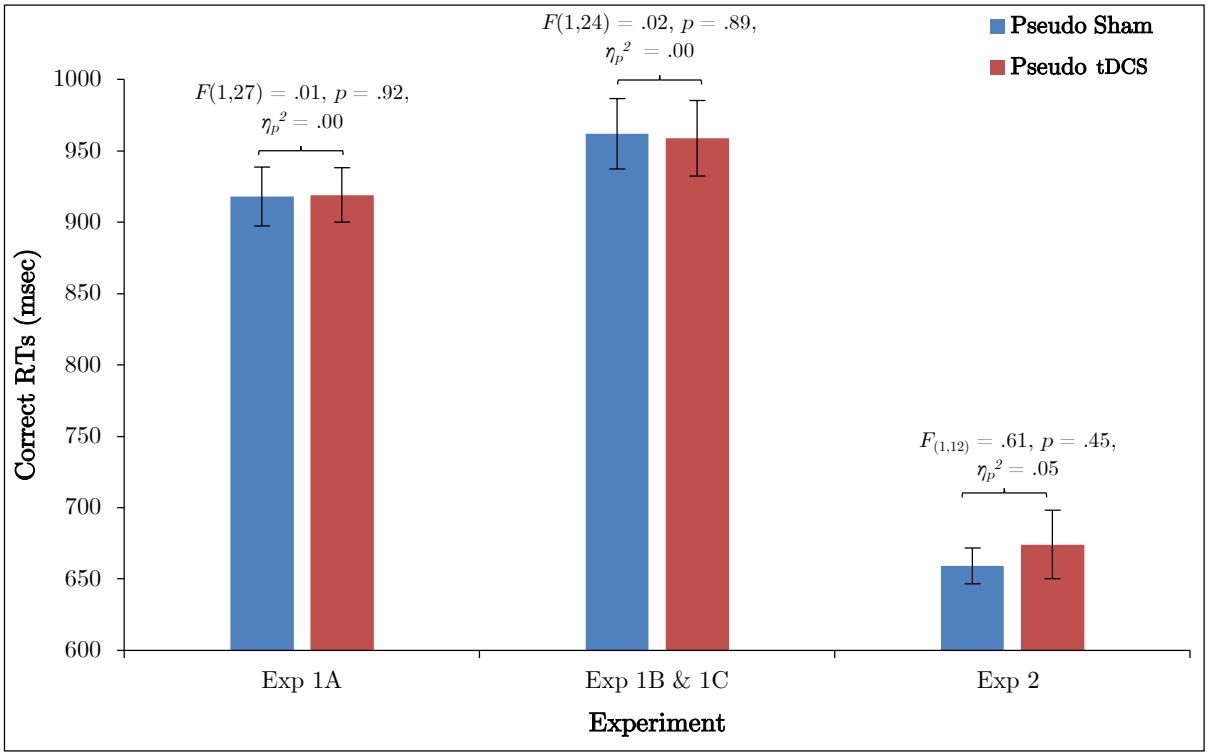
EFFECTS OF SESSION/STIMULATION

A. PICTURE NAMING - RTs

Experimental Groups

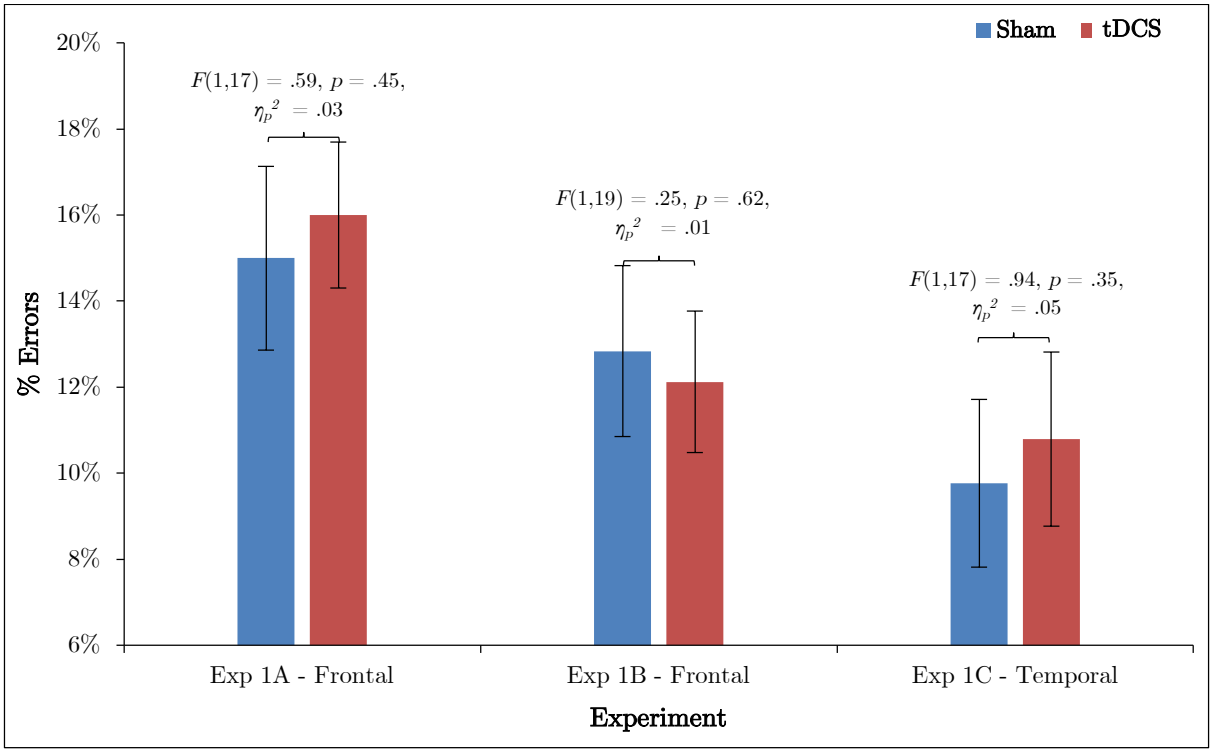


Control Groups

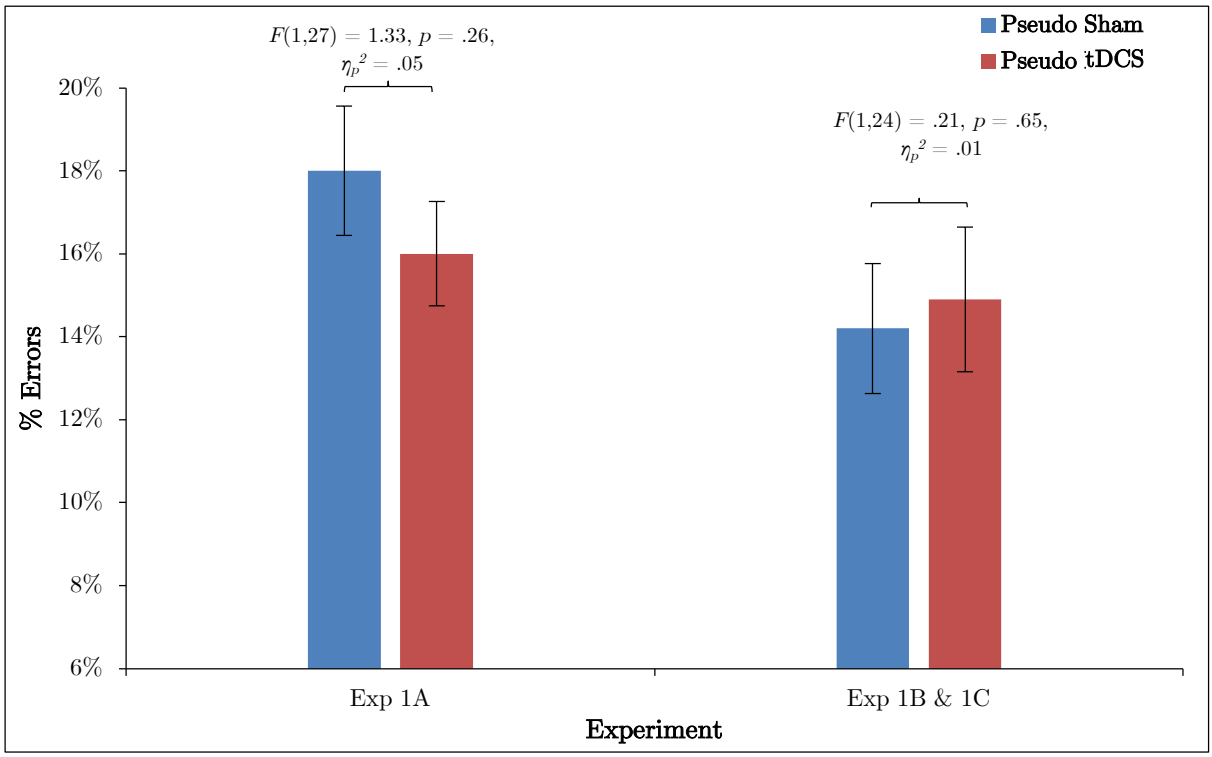


B. PICTURE NAMING - % ERRORS

Experimental Groups

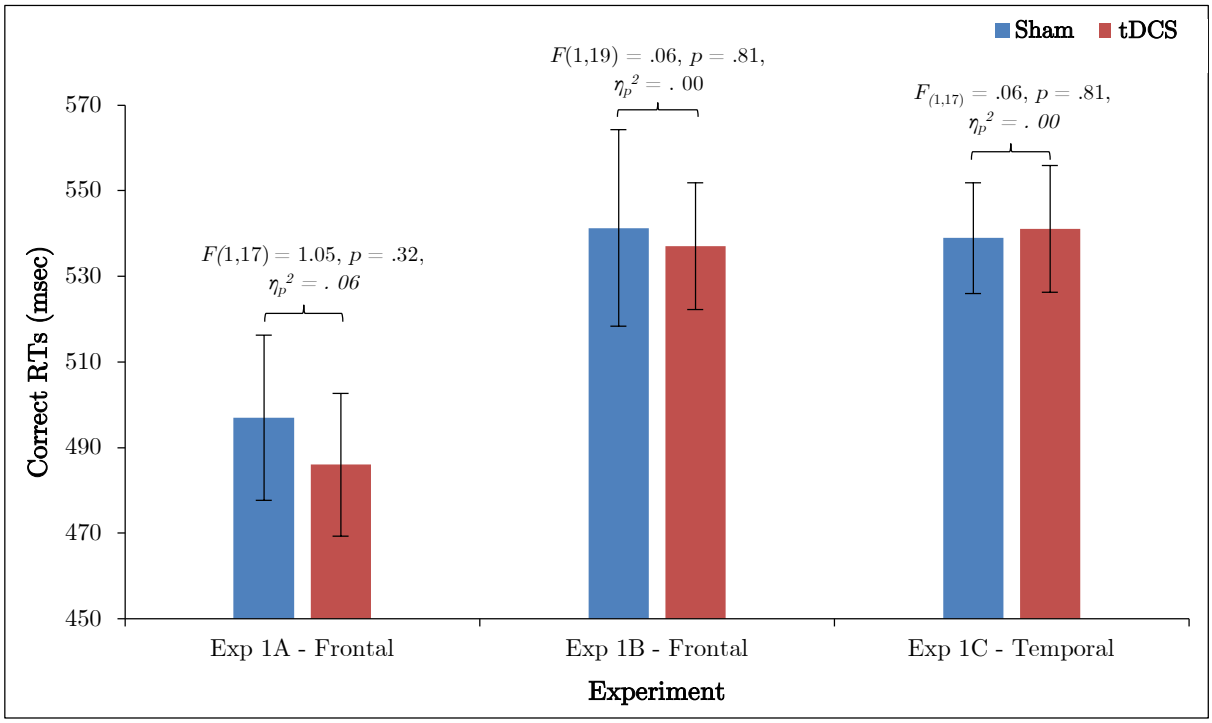


Control Groups



C. WORD READING - RTs

Experimental Groups



Control Group

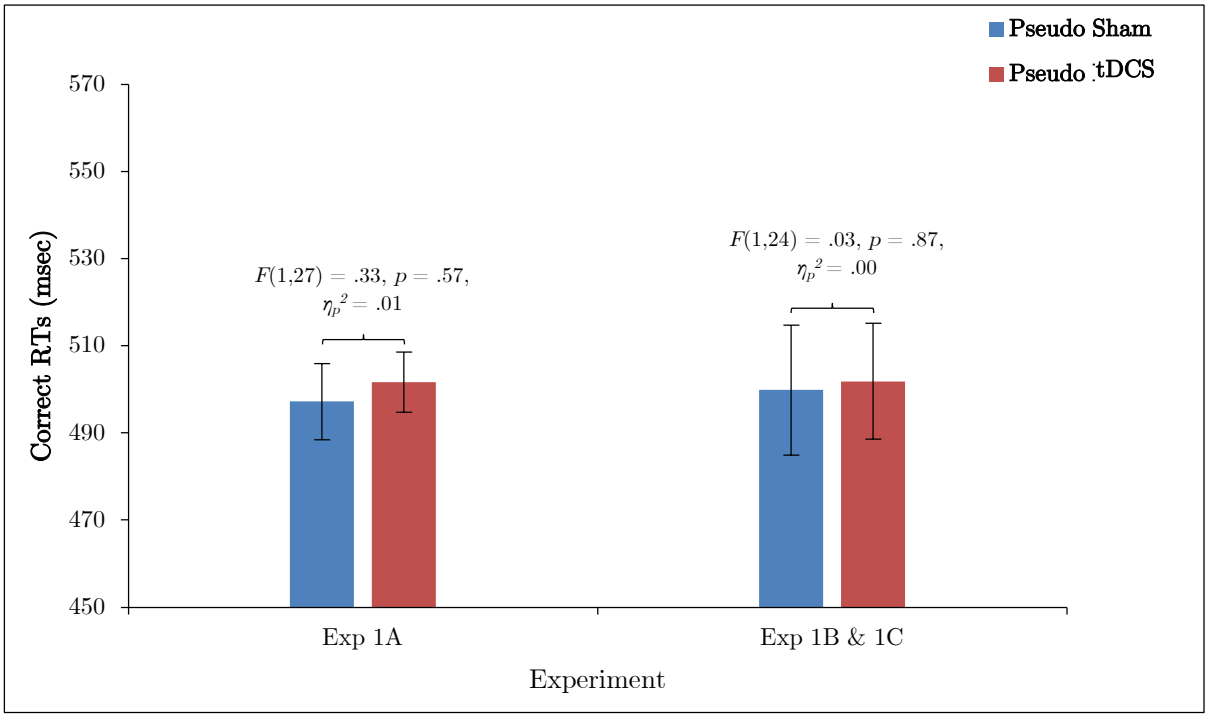


Fig 2.1 – Average correct RTs and accuracy differences between stimulation conditions (Sham vs tDCS for experimental group; Pseudo Sham vs Pseudo tDCS for control groups across experiments) across experiments. Error Bars indicate Standard Error.

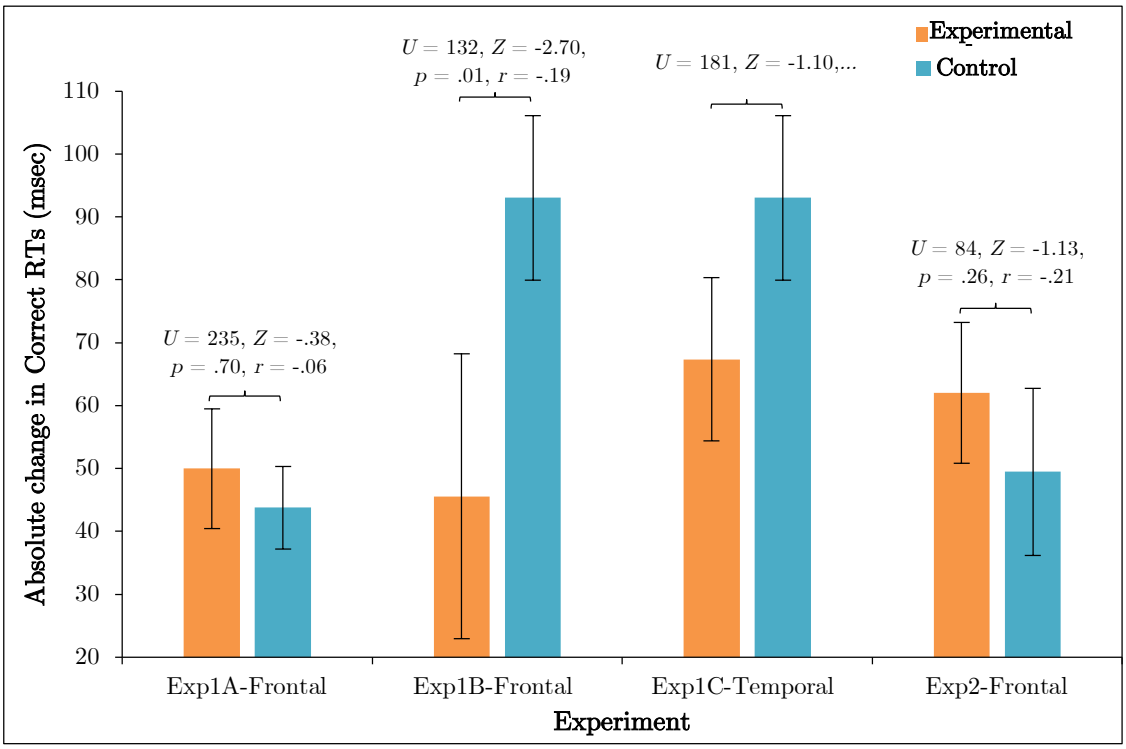
2.6.2.2 Direction-neutral effects of stimulation

Here, we considered tDCS effects when allowing for possible opposite outcomes across participants. We found that both participant groups were equally likely to improve or worsen performance relative to sham (or pseudo-sham), with both picture naming RTs (improve:worsen: $37{:}29_{\text{control}}$ vs $35{:}38_{\text{experimental}}$; $X(1) = .34$, $p = .34$), errors ($30{:}23_{\text{control}}$; $29{:}27_{\text{experimental}}$; $X(1) = .26$, $p = .61$), and reading RTs ($22{:}31_{\text{control}}$; $31{:}25_{\text{experimental}}$; $X(1) = 2.09$, $p = .15$).

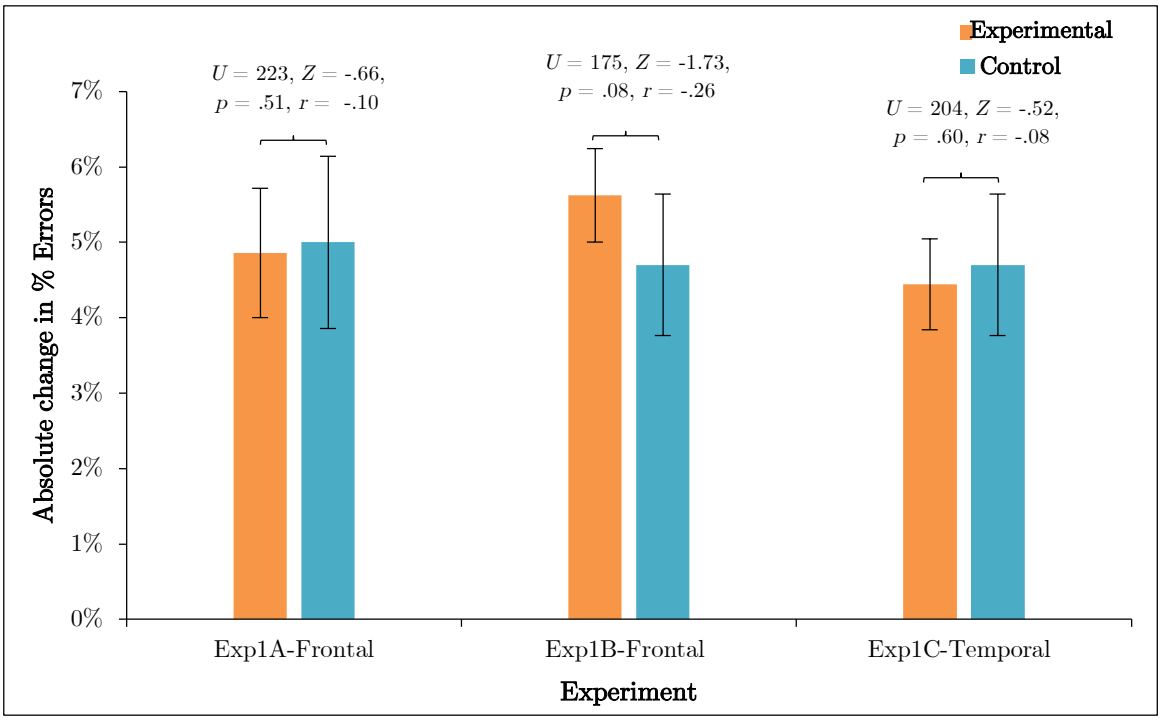
We also compared absolute differences between conditions in the experimental and control group via a series of Mann-Whitney U tests (as values were non-normally distributed). Results are shown in Figure 2.2. Overall, for picture naming RTs, the difference between conditions was smaller in the experimental group relative to the control group ($M \pm SE$: 56 ± 6 vs 64 ± 7 msec). This was the opposite of what was expected. It could be that stimulation (both tDCS and sham) reduces variability by increasing arousal and/or motivation. It has to be noted, however, that this effect was inconsistent with naming errors (5 ± .4 vs 5 ± 1%) and reading RTs (37 ± 5 vs 36 ± 5 msec).

ABSOLUTE DIFFERENCE OF SESSION/STIMUATION

A. PICTURE NAMING - RTs



B. PICTURE NAMING - % ERRORS



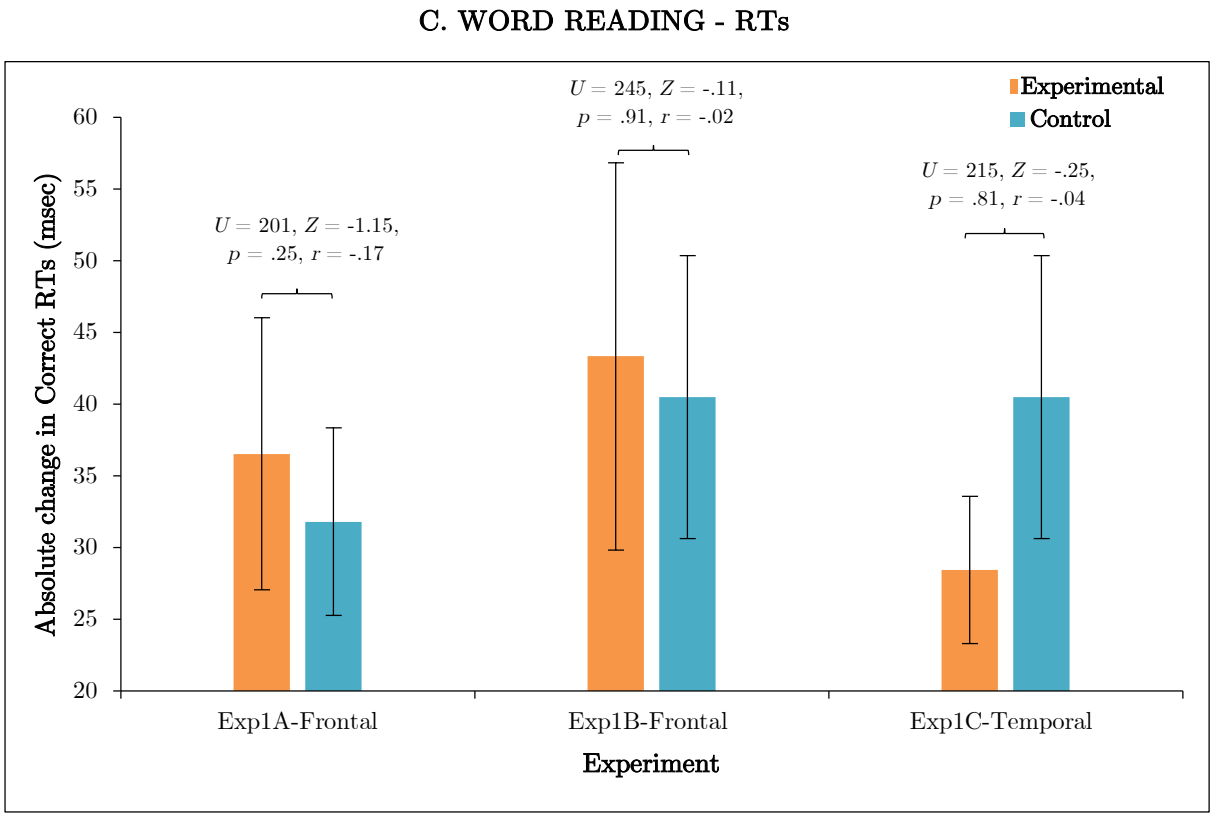


Fig. 2.2 – Absolute average correct RTs and accuracy differences between stimulation conditions (Sham vs tDCS for experimental group; Pseudo Sham vs Pseudo tDCS for control groups) across experiments. Error Bars indicate Standard Error.

2.6.3 Effects of tDCS on semantic interference

2.6.3.1 Cumulative interference

Performance across ordinal positions within sets of related items are shown in Figure 2.3. Across participant groups, tasks and conditions, our behavioural manipulation worked well. Picture naming shows a steady increase in latencies across positions; errors also show an increasing trend or no effect. Reading shows no systematic effect of position. Crucially, however, there are no detectable effects of tDCS – i.e., the increase in RTs with ordinal position was equivalent with or without tDCS. Numerically, performance was faster in real tDCS than sham in reading Experiment 1A (with a slight increase across positions similar to picture naming), but this difference is not significant (see below) and the opposite of what was seen in Experiment 1B.

We carried out separate repeated factor ANOVAs for each task, experiment and participant group, with *Ordinal Positions* (1-5) and *Condition* (*tDCS vs* Real for experimental group; *Pseudo-Sham vs Pseudo-tDCS* for control group) as within participant factors. With picture naming RTs, there was a main effect of *Ordinal Position*, with latencies increasing with each position (Experimental group: ${}_{1A}F(4,68) = 27.70$, $p < .001$, $\eta_p^2 = .62$; ${}_{1B}F(4,76) = 13.83$, $p < .001$, $\eta_p^2 = .42$; ${}_{1C}F(4,68) = 5.27$, $p = .001$, $\eta_p^2 = .24$; Control group: ${}_{1A}F(4,68) = 20.62$, $p < .001$, $\eta_p^2 = .43$; ${}_{1B/C}F(4,96) = 9.4$, $p < .001$, $\eta_p^2 = .28$). There was no main effect of Position with errors (Experimental group: ${}_{1A}F(4,68) = 1.69$, $p = .16$, $\eta_p^2 = .09$; ${}_{1B}F(4,76) = .84$, $p = .51$, $\eta_p^2 = .04$; ${}_{1C}F(4,68) = .45$, $p = .77$, $\eta_p^2 = .03$; Control group: ${}_{1B/C}F(4,96) = 1.3$, $p = .29$, $\eta_p^2 = .05$) except in the control group for Experiment 1A ($F(4,68) = 3.12$, $p = .02$, $\eta_p^2 = .10$). In this case, error rates increased after position three. With reading RTs, there was also a main effect of *Ordinal Position* in Experiment1A (Experimental group: $F(4,68) = 2.47$, $p = .05$, $\eta_p^2 = .13$; Control group: $F(4,108) = 3.4$, $p = .01$, $\eta_p^2 = .11$), but not in Experiment 1B or 1C (Experimental group: ${}_{1B}F(4,76) = .09$, $p = .99$, $\eta_p^2 = .01$; ${}_{1C}F(4,68) = .79$, $p = .54$, $\eta_p^2 = .04$; Control group: ${}_{1B/C}F(4,96) = .92$, $p = .46$, $\eta_p^2 = .04$). Crucially, there were no *Ordinal Position* x *Condition* interactions with naming RTs (Experimental group: ${}_{1A}F(4,68) = .75$, $p = .56$, $\eta_p^2 = .04$; ${}_{1B}F(4,76) = .51$, $p = .73$, $\eta_p^2 = .03$; ${}_{1C}F(4,68) = .34$, $p = .85$, $\eta_p^2 = .02$; Control group: ${}_{1A}F(4,68) = .38$, $p = .83$, $\eta_p^2 = .01$; ${}_{1B/C}F(4,96) = 1.1$, $p = .36$, $\eta_p^2 = .04$), naming errors (Experiment group: ${}_{1A}F(4,68) = .64$, p

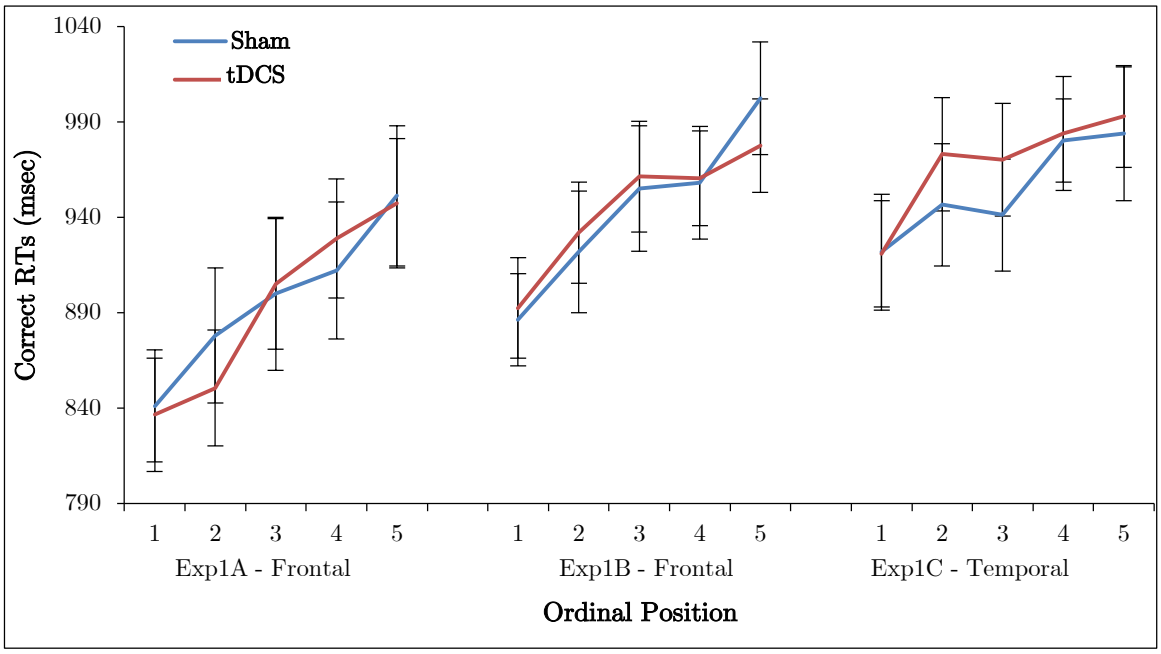
= .64, η_p^2 = .04; $_{1B}F(4,76)$ = .46, p = .76, η_p^2 = .02, $_{1C}F(4,68)$ = 1.13, p = .35, η_p^2 = .06; Control group: $_{1A}F(4,68)$ = .81, p = .52, η_p^2 = .03; $_{1B/C}F(4,96)$ = .63,p = .65, η_p^2 = .03), or reading RTs (Experimental group: $_{1A}F(4,68)$ = 1.1, p = .38, η_p^2 = .06; $_{1B}F(4,76)$ = .43, p = .78, η_p^2 = .02; $_{1C}F(4,68)$ = .71, p = .59, η_p^2 = .04; Control group: $_{1A}F(4,108)$ = .50, p = .74, η_p^2 = .02; $_{1B/C}F(4,96)$ = .14, p = .97, η_p^2 = .01).

A mixed factor ANOVA across all picture naming experiments with *Group* as a between-participant factor and *Ordinal Position* and *Condition* as within-participant factors showed no three way interaction between *Group* x *Condition* x *Ordinal Position* (naming RTs: $F(4,428)$ = .45, p = .77, η_p^2 = .04; errors: $F(4,428)$ = .95, p = .43, η_p^2 = .01; reading RTs: $F(4,428)$ = .34, p = .85, η_p^2 = .00).

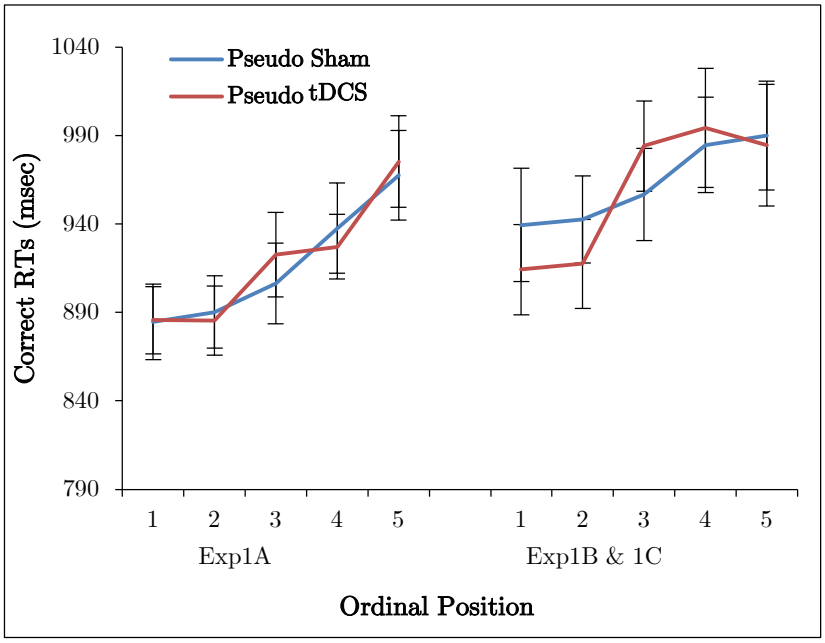
CUMULATIVE INTERFERENCE

A: PICTURE NAMING – RTs

Experimental Groups

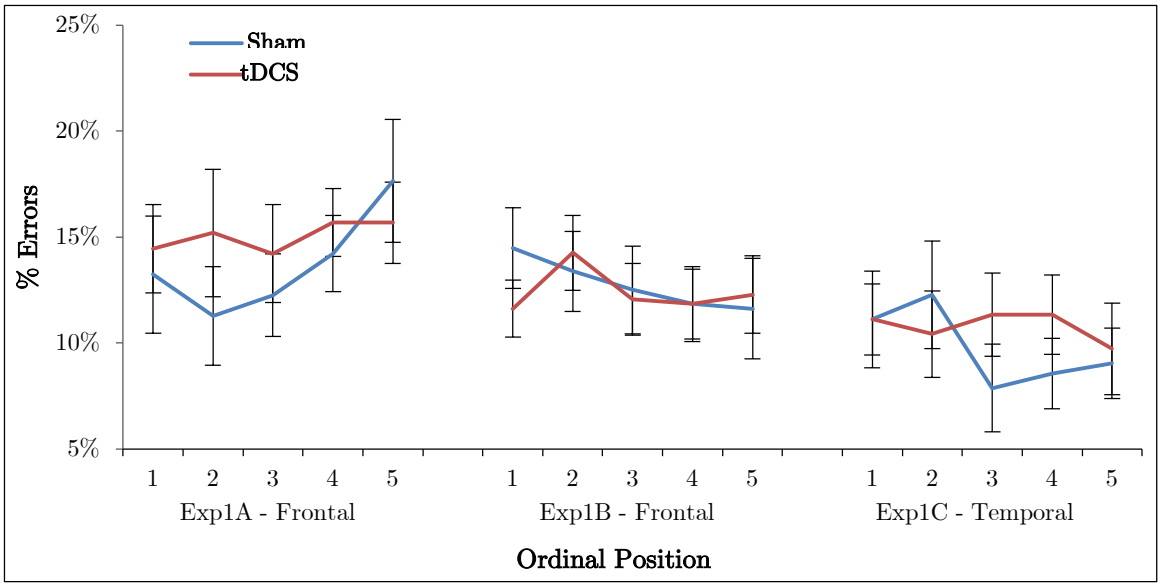


Control Groups

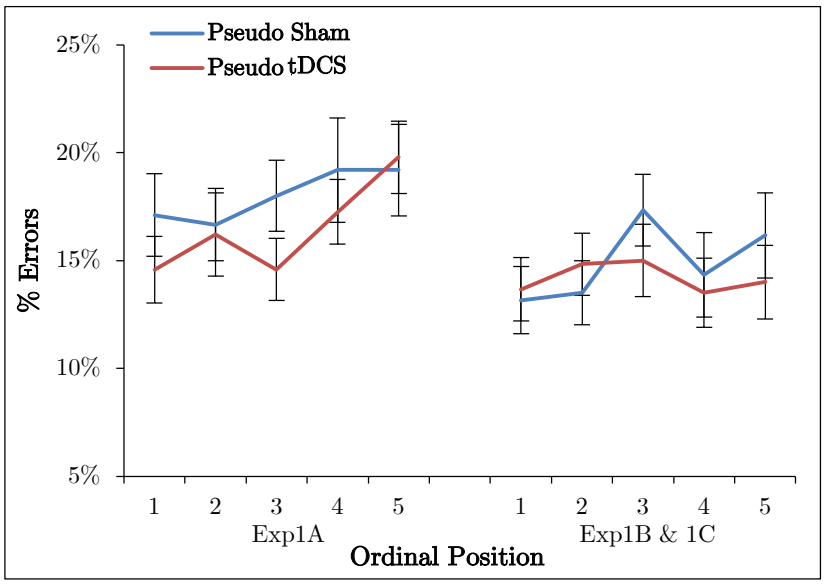


B. PICTURE NAMING – % ERRORS

Experimental Groups

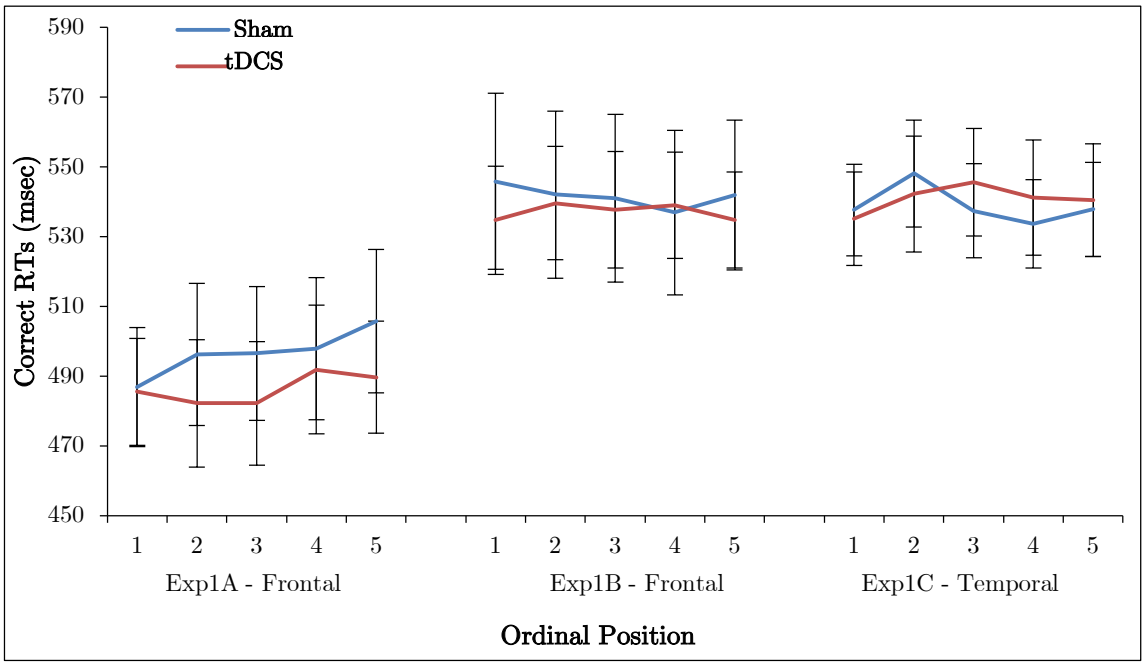


Control Groups



C. WORD READING – RTs

Experimental Groups



Control Groups

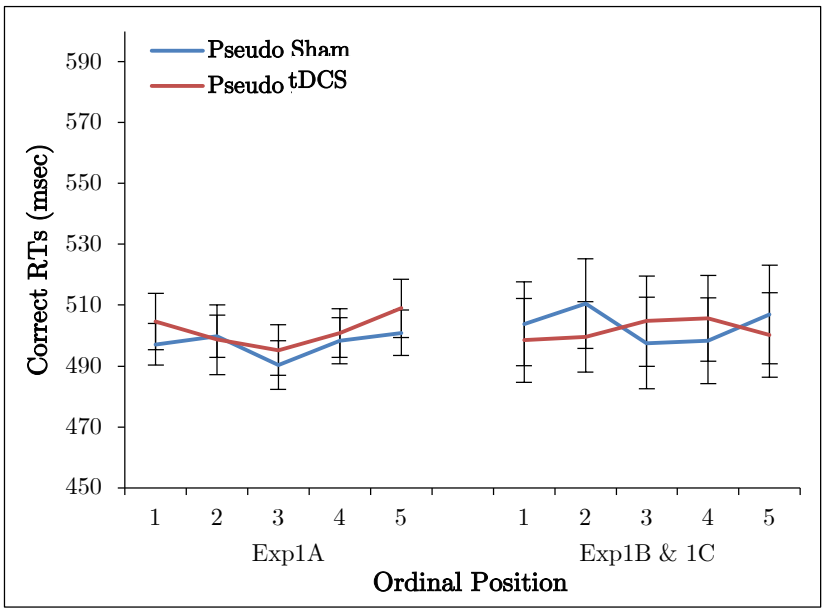


Fig 2.3 – Cumulative semantic interference effect. Average correct RTs and accuracy for ordinal positions across experiments and tasks. Error Bars indicate Standard Error.

2.6.3.2 Interference by relatedness and cycle

Results for Experiment 2 are shown in Figure 2.4. As expected, semantic relatedness interacted with cycle to modulate performance. For *unrelated* picture sets, participants became progressively faster with every repetition (or cycle), whilst, for *related* sets, naming latencies flattened after initial facilitation between the first and the second cycle. This pattern was produced by both the experimental and control group, and replicates what is typically found with this paradigm (Belke, 2013; Belke & Stielow, 2013).

We carried out a mixed factor ANOVA, with *Group* as a between-participants factor and *Relatedness*, *Cycle* and *Condition* (*tDCS vs Sham* for experimental group; *Pseudo-tDCS vs Pseudo-Sham* for control groups) as within-participants factors. There was a main effect of *Relatedness*, because related sets were slower than unrelated sets ($F(1,28) = 14.49$, $p = .001$, $\eta_p^2 = .62$), a main effect of *Cycle* ($F(3,84) = 45.90$, $p < .001$, $\eta_p^2 = .62$), because RTs became faster after the first cycle, and a significant interaction between *Relatedness* x *Cycle* ($F(3,84) = 28.12$, $p < .001$, $\eta_p^2 = .50$), because related sets were faster than unrelated sets in the first cycle, but then slower. Crucially, there was no main effect of *Group* ($F(1,28) = .06$, $p = .81$, $\eta_p^2 = .00$), nor a significant interactions2 between *Group* x *Condition* ($F(1,28) = .07$, $p = .79$, $\eta_p^2 = .00$), *Group* x *Condition* x *Relatedness* ($F(1,28) = .98$, $p = .33$, $\eta_p^2 = .03$), *Group* x *Condition* x *Cycle* ($F(1,28) = .85$, $p = .47$, $\eta_p^2 = .03$), and *Group* x *Condition* *Relatedness* x *Cycle* ($F(3,84) = 1.43$, $p = .24$, $\eta_p^2 = .05$).

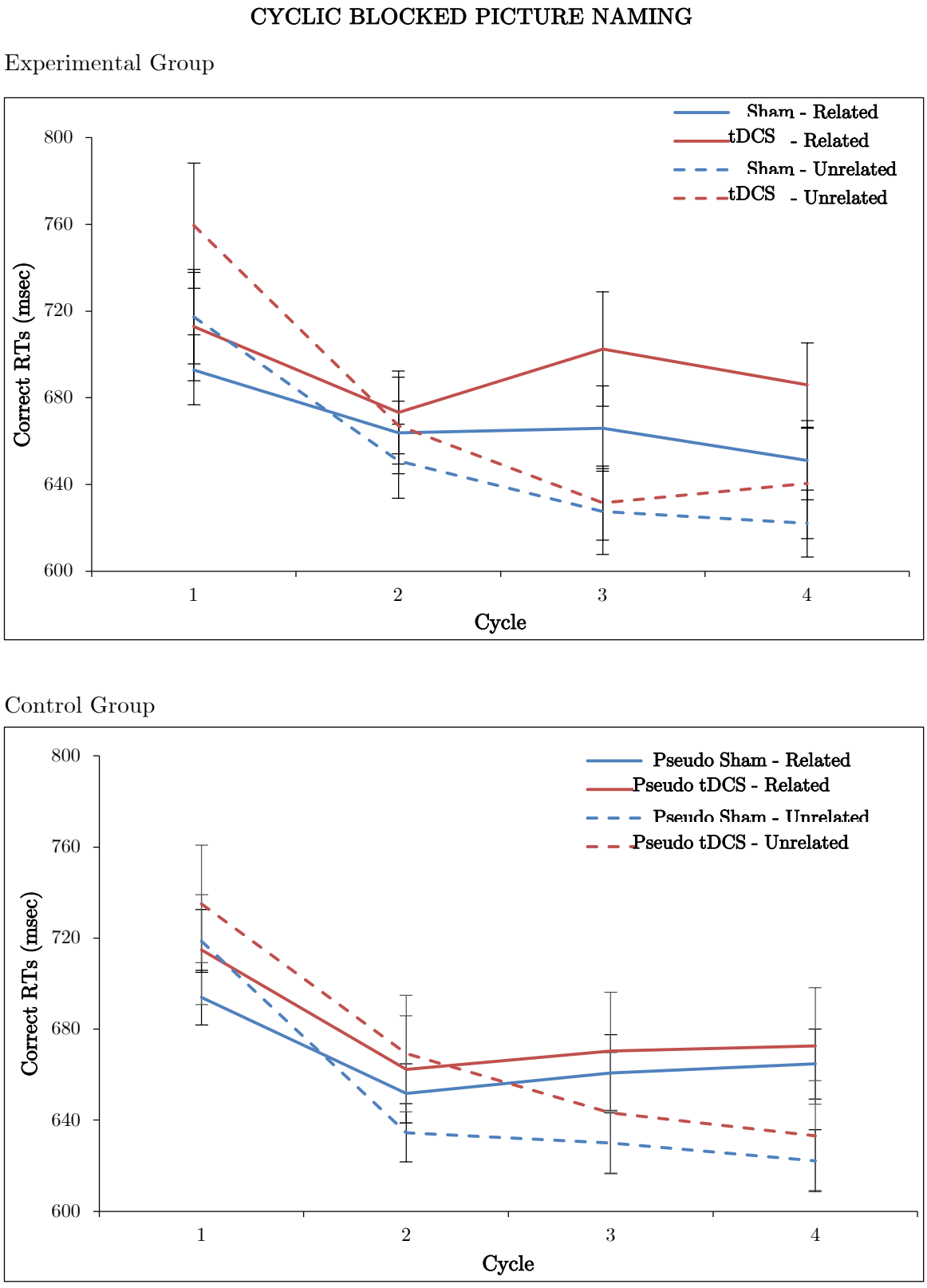


Fig 2.4 – Semantic interference effect by cycle. Average correct RTs for related and unrelated sets across cycles. Error Bars indicate Standard Error.

2.6.3.3 Aggregated interference

Here, we considered whether tDCS effects are detectable when interference effects are aggregated across conditions. For Experiment1A-C, we considered the difference in RTs between items in position 4 to 5 and items in position 1 to 2. For Experiment 2, we considered the difference between related and unrelated sets at cycle 4 (where the difference should be positive; with related sets being faster) and at cycle 1 (where the difference should be negative; with related sets being slower).

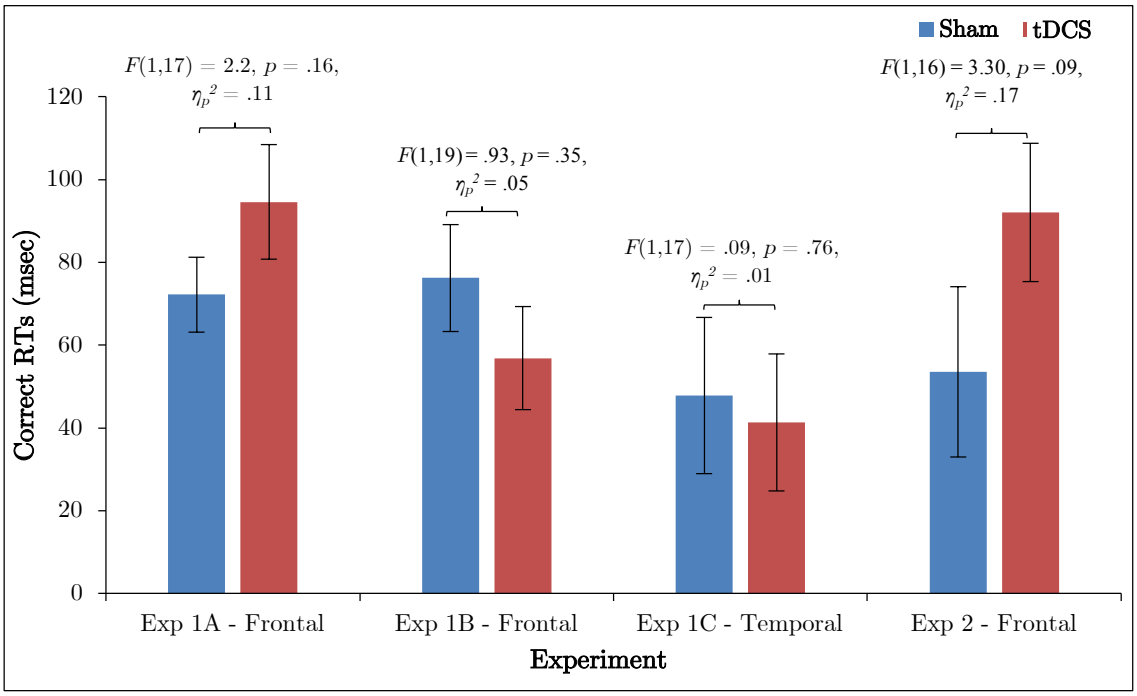
Aggregated interference effects across experiments, groups and conditions are presented in Figure 2.5. tDCS clearly had no consistent effect. In the experimental group, interference was larger with tDCS in Experiment 1A and 2, but the opposite was found in Experiment 1B and 1C. We carried out separate one-way ANOVAs for each experiment and participant group, with *aggregate interference* as a dependent measure and *Condition* as a within-participants measure. The results showed no significant main effect of *Condition* (Experimental group: $F < 3.30$, $p > .09$, $\eta_p^2 < .17$; Control group: $F < 1.04$, $p > .32$, $\eta_p^2 < .04$). We also carried out a mixed factor ANOVA with *Group* as a between-participants factor and *Condition* as a within participants factor. Crucially, there was no *Group* x *Condition* interaction ($F(1,137) = .01$, $p = .93$, $\eta_p^2 = .00$).

2.6.3.4 Interaction with cortical loci of stimulation

Given the possibility that tDCS could reduce a semantic interference effect with frontal stimulation, but increase it with temporal stimulation we carried out a mixed factor ANOVA with *aggregate interference* as a dependent measure, *Site* (Frontal-Stimulation-Exp 1B *vs* Temporal-Stimulation-Exp 1C) as a between-participants factor and *Condition* as a within-participants factor. Again, there was no main effect of *Condition* ($F(1,36) = .80$, $p = .38$, $\eta_p^2 = .022$), *Site* ($F(1,36) = 1.89$, $p = .18$, $\eta_p^2 = .05$) and no *Condition* x *Site* interaction ($F(1,36) = .21$, $p = .65$, $\eta_p^2 = .01$).

AGGREGATED SEMANTIC INTERFERENCE – RTs

Experimental Groups



Control Groups

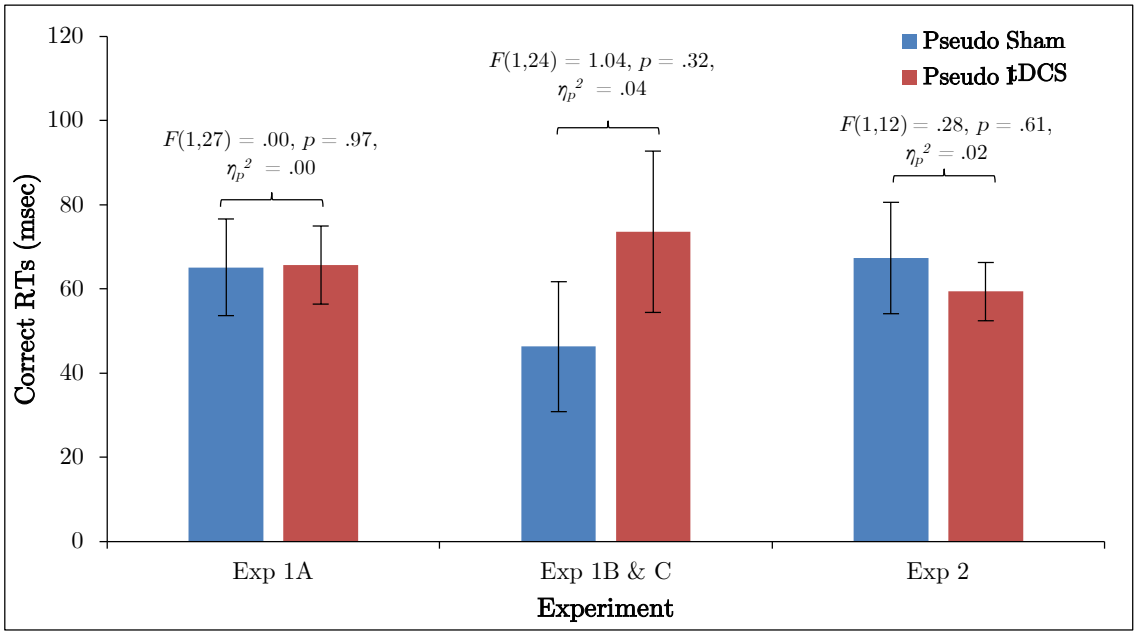


Fig 2.5 – Semantic interference effect averaged across conditions. For experiment 1, interference measured as the differences between the last two and first two ordinal positions; for experiment 2, interference measured as the difference between related and related blocks at cycle 4 versus cycle 1; e.g., (related-unrelated at cycle 4) minus (related-unrelated at cycle 1).

2.6.3.5 Direction-neutral effects of stimulation

Here, we compared absolute differences in interference across stimulation conditions in the experimental and control groups. Results are shown in Figure 2.6. Mann-Whitney *U* tests showed that interference effects changed more across conditions in the experimental than in the control group in Experiment 2, but not in any other experiment and effects were numerically in the opposite directions in Experiments 1B and 1C.

2.6.3.6 Effect of stimulation by magnitude of interference

To assess whether tDCS effects were dependent on the level of semantic interference we grouped experimental participants into those who showed high versus lower levels of semantic interference. We collapsed picture-naming data for all experiments and conducted a median split on the size of semantic interference across both the tDCS and sham conditions. Figure 2.7 shows that RTs across participants showing high versus low interference effects were not moderated by stimulation. A mixed factor ANOVA, with Interference (High vs Low) as a between-participants factor and Condition as a within-participant factor showed no significant *Interference* x *Condition* interaction ($F(1,71) = 1.27$, $p = .26$, $\eta_p^2 = .02$), suggesting that tDCS effects were not moderated by the size of the semantic interference effect.

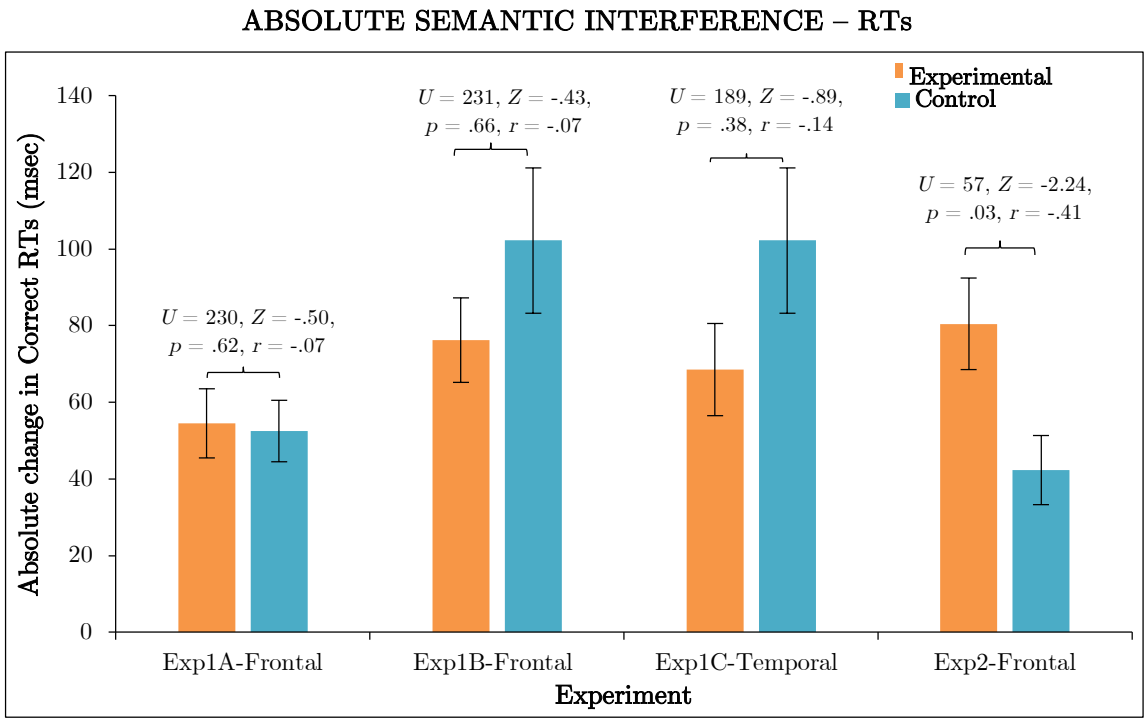


Fig 2.6 – Semantic interference effect in terms of absolute differences in RTs between stimulation conditions (Sham vs tDCS for the experimental group; Pseudo Sham vs Pseudo tDCS for control group) across experiments. Error Bars indicate Standard Error.

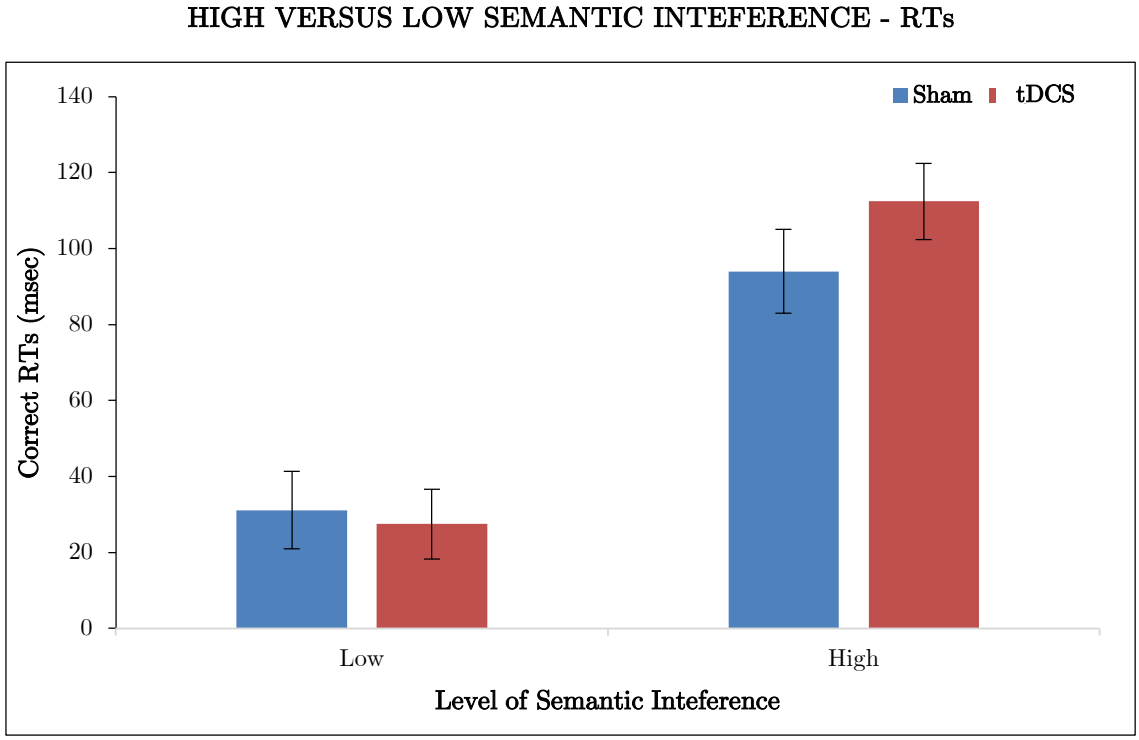


Fig 2.7 – High versus low semantic interference effects effect in terms of RTs across stimulation conditions (Sham vs tDCS for the experimental group; Pseudo Sham vs Pseudo Real for control group) and experiments. Error Bars indicate Standard Error.

2.6.1 Effects of stimulation by item difficulty

We assessed if tDCS effects were limited to items that recruited greater cognitive resources by running a so-called *Vincentisation* analysis. For each task (reading and picture naming), we ranked each participant's RTs within each ordinal position (Experiment 1) or Cycle (Experiment 2), and then placed the RTs into four bins according to speed (e.g., *very slow*, *slow*, *fast*, *very fast*), each with 25% of data. This was done separately for each condition (i.e., tDCS and Sham; Pseudo-tDCS and Pseudo-Sham). Results in Figure 2.8 show that conditions in the experimental and control groups did not systematically differ depending on speed bin.

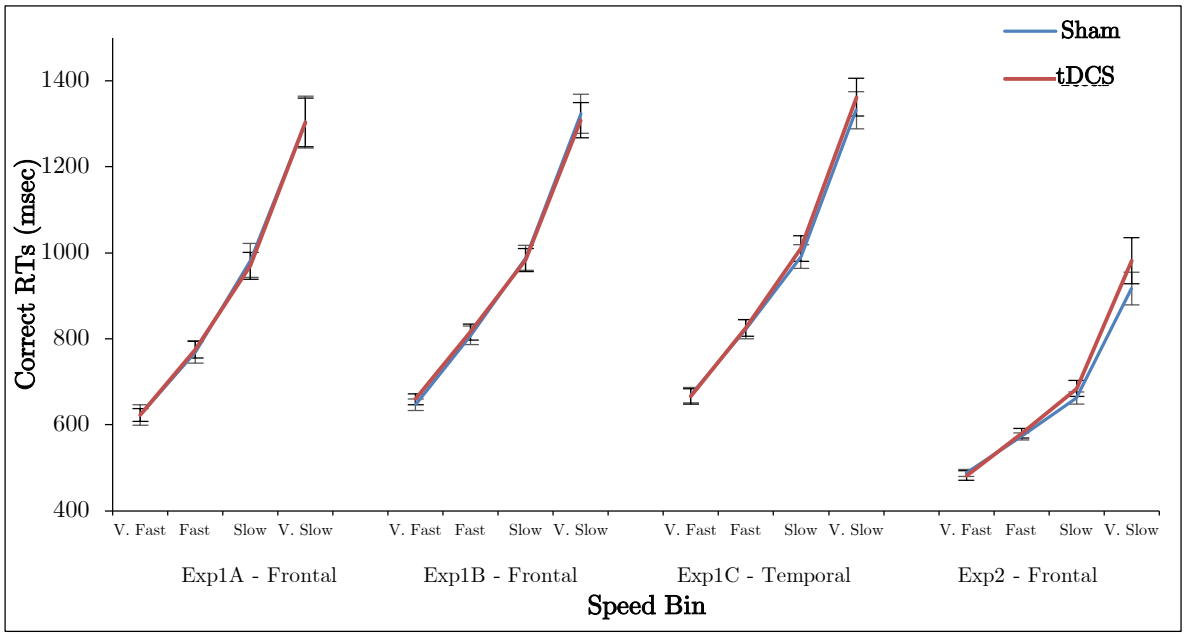
We carried out separate mixed factor ANOVAs for each experiment, with *Group* (*Experiment vs Control*) as a between participants factor and *Speed Bin* (1, 2, 3, 4) and *Condition* (Sham *vs* tDCS for the experimental group; Pseudo-Sham *vs* Pseudo-tDCS for control group) as within-participants factors. Effects of speed bins are expected and not of interest. Crucially, there was no significant *Speed* x *Bin* x *Group* x *Condition* interaction for picture naming RTs (${}_{1A}F(3,132) = .43$, $p = .74$, $h^2_p = .01$; ${}_{1B}F(3,129) = .14$, $p = .94$, $\eta_p^2 = .00$; ${}_{1C}F(3,123) = .78$, $p = .51$, $\eta_p^2 = .02$; ${}_{2}F(3,84) = .43$, $p = .74$, $\eta_p^2 = .02$) or reading RTs(${}_{1B}F(3,129) = 1.21$, $p = .31$, $\eta_p^2 = .03$), except for Experiment 1A and 1C (${}_{1A}F(3,132) = 3.32$, $p = .02$, $\eta_p^2 = .07$; ${}_{1C}F(3,123) = 2.68$, $p = .05$, $\eta_p^2 = .06$).

We carried out separate repeated factor ANOVAs to unpack the three-way interaction found in reading RTs for Experiment 1A and 1C, focusing on experimental participants only. We found no significant *Speed* x *Bin* x *Condition* interaction (${}_{1A}F(3,51) = 2.18$, $p = .10$, $\eta_p^2 = .11$; ${}_{1C}F(3,51) = 1.64$, $p = .19$, $\eta_p^2 = .09$). Thus, overall, the data showed that tDCS did not selectively modulate performance under high cognitive load.

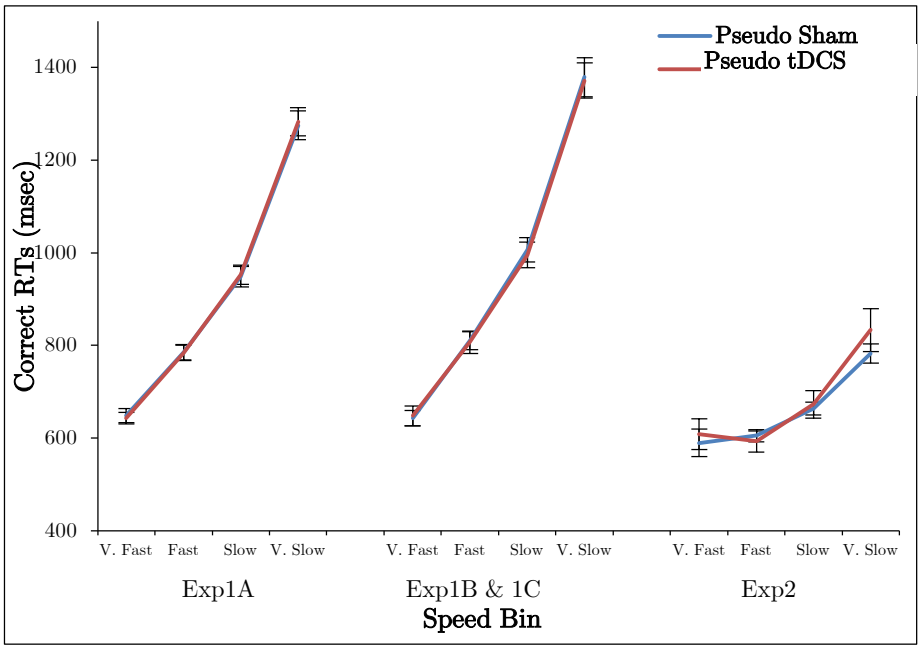
EFFECT OF STIMULATION ACROSS SPEED BINS

A. PICTURE NAMING – RTs

Experimental Group

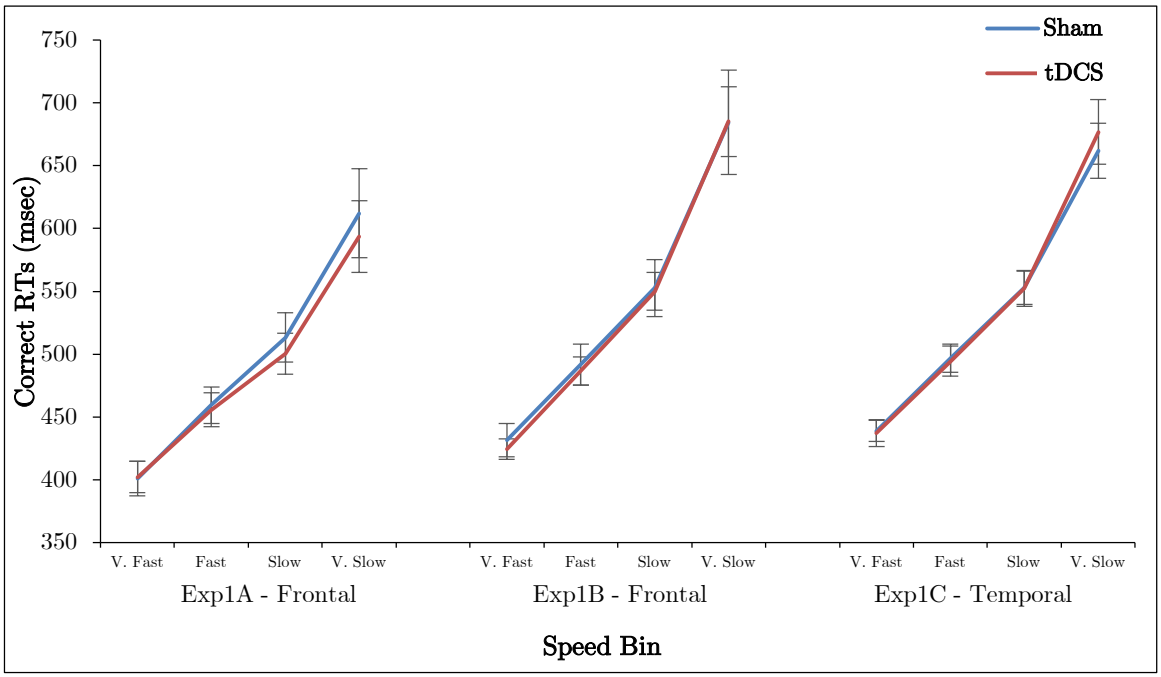


Control Group



B: WORD READING – RTs

Experimental Group



Control Group

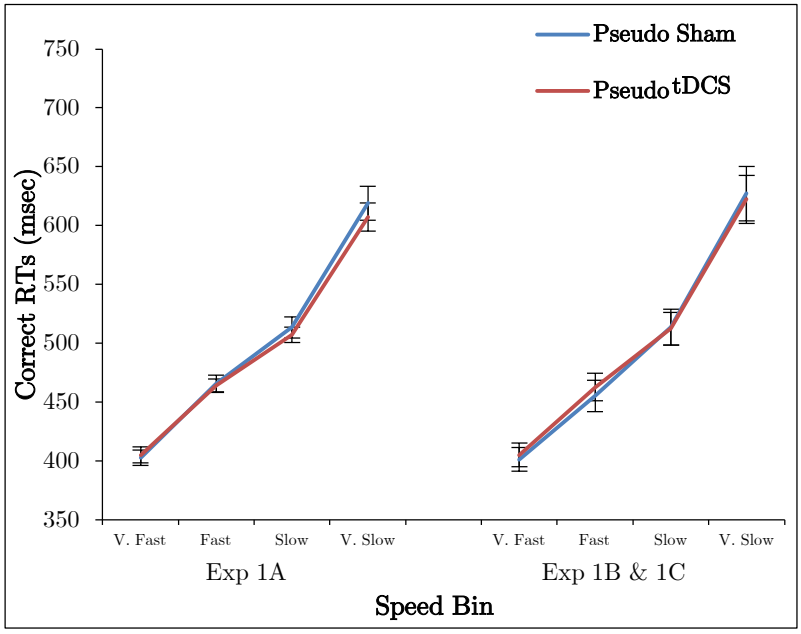


Fig 2.8 – Average correct RTs following Vincentisation. Average RT across speed bins, experiments, and participant groups. Error Bars indicate Standard Error.

2.7 Discussion

In the Introduction, we outlined how recent reviews have reported effects of tDCS to be small, inconsistent and not significant when averaged across studies (e.g., Horvath et al., 2015a). Our experimental investigation aimed to provide further evidence for whether tDCS can modulate language processing in normal healthy participants. We carried out four studies with different groups of participants which employed tasks typically used to probe lexical access and word production – namely picture naming and word reading – and used stimulation protocols typically used by studies reporting positive effects (e.g., 1-1.5 mA of anodal stimulation to frontal and temporal areas for 15-25 min during task performance). We made particular efforts to assess whether potential null effects could be masked by variability in the net outcome of tDCS depending on individual baseline levels of cortical excitability and task requirements. We maximized our chances of demonstrating a possible reversal of the advantages generally predicted for language tasks with anodal tDCS of left-hemisphere areas by: 1. Considering task conditions affording a high level of competition from semantically related items, that is, comparing tDCS effects on sets of related versus unrelated items; 2. Considering individual variability in the net outcome of tDCS, that is, assessing whether, with the same task, some participants may show significant facilitation and others significant worsening of performance; 3. Contrasting activation of different areas with the hypothesis that frontal stimulation may boost selection mechanisms, thus reducing interference, while temporal activation may boost lexical activation, thus, increasing interference; 4. Considering preferential effects for participants who demonstrated high semantic interference; 5. Considering possible enhanced/reduced effects of tDCS on difficult to name items. Despite our best efforts, we found no evidence of performance modulation due to the tDCS.

Our results contribute to growing doubts surrounding the reliability of tDCS applied within one stimulation session as a tool to modulate cognition in populations of neurologically intact participants. The effects of tDCS on semantic interference are particularly representative. With temporal stimulation one study found *reduced* interference (Meinzer et al., 2016), one found enhanced interference (Pisoni et al., 2012) and two found no effect (our own and Henseler et al., 2014). With frontal stimulation three studies found reduced

interference (Meinzer et al., 2016; Pisoni et al., 2012; Wirth et al., 2011), but two others found no effect with the same paradigm (our own study) or with a different paradigm (Henseler et al., 2014). Why these differences? A close consideration of the tDCS paradigms employed by these studies does not reveal any clear difference which may be responsible for different outcomes. The three studies which found a reduction of interference effects after frontal stimulation used parameters in the range covered by our experiments. Like us, they stimulated the left inferior frontal gyrus; placed the electrode on the contralateral supraorbital area; used a current density in a similar range (mA/cm^2 of .029, .057, .080; ours .11 to .06); a similar size of the reference electrode 35 to 100 cm^2 (our 35 cm^2), a similar size of active electrode (25 to 35 cm^2; our 9 to 25 cm^2) and administered the current for a similar duration (20 to 25 min; ours 15 to 25 min). Of course, one may always argue that we did not use the right combination of parameters. However, lack of empirical evidence in addition to lack of any appropriate mechanistic model that can provide specific predictions means that we are in the dark when searching for the right parameter combination (for a discussion, see de Berker et al., 2013; Horvath et al., 2016).

Another possible explanation for our null effects is of course lack of power. Our total samples of 56 and 73 participants for reading and naming respectively allowed us good power to detect medium (.5) or strong (.8) effects of tDCS ($1\text{-}\beta > .96$) for both. However, the power to detect a *small* effect of tDCS (*effect size* = .25, $\alpha = .05$) was limited even within a within-participants design like ours ($1\text{-}\beta = .45$ and .56 for reading and naming). To prove or disprove a *small* effect of tDCS with strong statistical power would have required a sample of 128 participants (*effect size* = .25, $1\text{-}\beta = .8$, $\alpha = .05$). This is inconsistent with standards in the field. Most published studies report samples between 10 and 25 participants (see Horvath et al., 2015a; Price et al., 2015; Tremblay et al., 2014). One may want to encourage studies with many more participants, but the fact remains that if effects of tDCS are so small, tDCS is not a tool fit for purpose in the way it is currently employed for modulation of normal cognition. Meta-analyses are of course one way to tackle the issue of small sample sizes. In a review of studies assessing effects of tDCS in reading and picture naming, we pooled studies using a similar protocol to the present study – i.e., applied left anodal tDCS to frontal/temporal lobes – and included the present study. This gave a total sample size of

roughly 200 participants. Even with this sample size, we found no evidence of a tDCS effect (see Westwood & Romani, 2017[6]).

It is possible that future studies will elucidate conditions where single session tDCS is efficacious even in healthy participants. It is also possible, however, that cortical excitability in healthy brains is already close enough to an optimal level that cannot be bettered and/or that homeostatic mechanisms come into play to reduce excessive levels of activation, thus, nullifying any effect of tDCS (Krause & Cohen Kadosh, 2014). Instead, effects of tDCS may only be reliable in neurologically damaged participants where targeted regions may have a pathologically reduced level of excitability (for a review, see Silvanto et al., 2008). A recent review of extant literature on post-stroke aphasia composed of twelve studies (de Aguiar et al., 2015) indicated a general benefit of tDCS across language tasks and types of therapy with varied stimulation protocols. The results showing improvements in picture naming are particularly relevant here (see Fiori et al., 2011; Floel et al., 2011; Kang, Kim, Sohn, Cohen, & Paik, 2011; Lee, Cheon, Yoon, Chang, & Kim, 2013; Marangolo et al., 2013; Saidmanesh, Pouretemad, Amini, Nilipor, & Ekhtiari, 2012; but see also Monti et al., 2008).

Alternatively, positive results may be dependent on dose of stimulation (see Meinzer et al., 2014). Positive results with aphasic participants are obtained when tDCS is administered in conjunction with naming once or twice a week for a number of weeks (sessions ranging from 5 to 10). It is possible, therefore, that the key for positive effects of tDCS is not whether the treated population is healthy or impaired, but the stimulation dose and/or repeated application across a number of sessions. It is also possible that positive effects are more likely in tasks that require novel cognitive operations, which are less established in the brain, such as during the acquisition of new processes or representations. Novel operations may be easier to manipulate than operations already well established, such as naming common items (for a similar argument, see Jacobson et al., 2012). It has been shown that tDCS can modify synaptic plasticity by modulating levels of glutamate, GABA, and other neurotransmitters (e.g., dopamine, serotonin, acetylcholine; for extensive reviews, see Medeiros et al., 2012;

[6]See Chapter 3

Stagg & Nitsche, 2011). This may permit modulation of learning. Indeed, a number of studies have shown enhanced learning following repeated stimulation even in normal participants (Kadosh, Soskic, Iuculano, Kanai, & Walsh, 2010; Dockery, Hueckel-Weng, Birbaumer, & Plewnia, 2009; Meinzer et al., 2014; Reis et al., 2009). Flöel et al. (2008) reported enhanced novel word learning even after a single stimulation session, although the effect vanished after one week.

2.7.1 Conclusions

The bias to publish significant results combined with a lack of appetite for replication (see, Open Science Collaboration, 2015; Vannorsdall et al., 2016), may have given the research community a false sense of tDCS effectiveness. Our results suggest that the unreliability of tDCS results should be taken as a starting point and as a challenge that needs addressing, rather than assuming a level of a reliability that is not there. Across a variety of conditions and analyses, we found no evidence that online tDCS could modulate word retrieval in healthy participants. We performed analyses which considered possible causes of variability, but found no significant results. Further studies should expand on these analyses. Further studies should also assess whether positive effects can be obtained even in healthy participants when stimulation is carried out across different sessions and/or when it involves learning of novel words rather than the modulation of a consolidated vocabulary as in the present study. More generally, our results suggest that the efficacy of tDCS to modulate normal cognition needs to be carefully re-evaluated.

Chapter 3: A meta-analysis evaluating the effects of tDCS on picture naming and word reading

3.1 Overview

In this chapter, the null effects reported in Chapter 2 led me to evaluate the foundational claim that anodal tDCS can modulate picture naming and word reading in healthy participants in a single session. I therefore carried out a meta-analysis to assess the overall effectiveness of tDCS, pooling studies by different stimulation protocols and outcome measures to identify conditions in which tDCS may be more or less effective. I submit below the final published version (Westwood & Romani, 2017).

3.2 Introduction

Transcranial direct current stimulation (or tDCS) is a popular technique used to modulate cortical excitability via a weak electric current applied on the scalp. The technique is used widely across studies aiming to enhance cognitive functions, with its popularity rising sharply in recent years. According to PubMed, only a few dozen papers were published in the early 2000s, but several thousand have been published in the past ten years, many of which report positive gains on a variety of cognitive tasks. However, a growing number of researchers are calling for the re-evaluation of tDCS in healthy samples because of weak and inconsistent effects (see Horvath et al., 2015a; Underwood, 2016; Walsh, 2013) and broader concerns about the reproducibility of results in neuroscience (see Open Collaboration, 2015; see also Cumming, 2013; Ioannidis, 2005). Here, we carried out several meta- analyses to assess whether single sessions of tDCS can reliably modify performance on language tasks in healthy participants, an area which has received less attention by previous reviews.

Horvath et al. (2015a) were the first to conduct a quantitative review which indicated little – if any – evidence of significant cognitive effects with single sessions of tDCS in healthy participants. Null effects were reported across polarities (anodal or cathodal), cognitive domains (executive functions, language, visual and verbal memory, and miscellaneous higher-cognitive functions), and stimulated areas in the left and right hemisphere (e.g.,

frontal, temporal, motor and parietal regions). In a second quantitative review of neurophysiological effects (Horvath et al., 2015b), tDCS was only effective in modifying motor evoked potentials (MEPs or muscular 'twitches'). However, these re- views have been criticised for their restrictive inclusion criteria (see Price & Hamilton, 2015). For the cognitive review, to reduce the effects of idiosyncratic stimulation protocols, the authors excluded out- come measures that were not reported by two or more separate labs, which narrowed the number of eligible studies. Unfortunately, this meant many analyses – particularly those including language experiments – pooled just two or three studies to fit the varying degrees of stringency for each analysis (e.g., by outcome measure, by polarity, by stimulation timing). Nonetheless, across all 59 analyses, of which 12 pooled more than 5 studies, no significant results in favour of tDCS were found.

In a meta-analysis, Jacobson et al. (2012) attempted to verify the assumption that anodal (or excitatory) tDCS versus cathodal (or inhibitory) tDCS leads to respective improvement versus impairment in performance - an assumption that underpins nearly all cognitive studies using tDCS. The authors found that 81% and 47% of cognitive studies (n = 34) showed, respectively, the expected anodal related improvement and cathodal related impairment across a variety of tasks, including attention, working memory and language. However, this review pooled data from healthy and patient samples. Moreover, the authors were interested in the reliability of outcomes with anodal and cathodal tDCS, rather than an effect of tDCS per se. Therefore, they excluded null results, results that contradicted the anodal-enhancement versus cathodal-impairment assumption, and reported only the largest effects when multiple effects were reported by a study. This meant that of the 4 included studies measuring effects of tDCS on language production with healthy volunteers, 3 reported positive effects of tDCS. This, however, masked the fact that 26 negative effects were excluded.

Other meta-analyses with healthy participants focusing largely on working memory/short-term memory tasks (WM/STM) reported equally weak and/or inconsistent effects. Hill et al. (2016) found small but positive effects on both accuracy and reaction times across n-back, span and Sternberg tasks, while other studies found positive effects only in reaction times (Brunoni & Vanderhasselt, 2014; Dedoncker et al., 2016a). Importantly, a

comprehensive review by Mancuso et al. (2016) found positive results were limited to studies using training paradigms – e.g., where performance on a WM task (e.g., n-back) was assessed after practicing the same task or a different WM task (e.g., Sternberg) under stimulation.

We know of only one published review that has focused on an effect of tDCS on language tasks in healthy participants. Price et al. (2015) reviewed studies involving verbal fluency (n = 6) and word learning (n = 2). Anodal tDCS improved accuracy scores significantly when pooling: a) all studies together; b) four studies where tDCS was applied prior to the task (i.e., offline stimulation) or c) three studies measuring offline tDCS with verbal fluency. However, effects were small to moderate in size (roughly ~ 0.5), and significant outcomes appeared to rely on the exceptionally large effect sizes from one study measuring fluency (~ 1.2, Cattaneo et al., 2011) and one measuring word learning (~ 0.8, Flöel et al., 2008). Furthermore, what is worse, one of these studies has proven difficult to replicate (i.e., Cattaneo et al., 2011; see Penolazzi et al., 2013a; Vannorsdall et al., 2016; but also Cattaneo et al., 2016 for response). Horvath (2015) also pointed out that the offline effect for fluency tasks would become non-significant if some data from studies excluded by the authors were instead included and if some mistakes in effect sizes estimates were corrected.

Our review will examine effects of tDCS in picture naming and word reading. Several reasons have informed our choice.

1. Reading and naming are widely considered to be good indicators of language competence. Moreover, although these tasks require different levels of processing (semantic, phonological/orthographic, articulatory) there is strong consensus that all these processes are based on relatively limited, frontal and temporal regions in the left hemisphere, which gives us confidence for what stimulation sites to focus on (see Indefrey, 2011; Indefrey and Levelt, 2004; Price, 2000).

2. These tasks tap resources such as semantic memory, executive functions, and working memory which are used beyond language tasks making naming and reading good proxies for the general effectiveness of tDCS for other cognitive functions (Badre & D'Esposito, 2009; Binder, Desai, Graves & Conant, 2009).

3. Studies using naming and reading have reported significant effects of tDCS, but consistent effects have been limited to individuals with language impairments following a stroke. For example, tDCS has been found to facilitate speech and language therapy for word finding difficulties in aphasic patients (see Cappon et al., 2016; de Aguiar et al., 2015; Monti et al., 2013; Crinion, 2016; Elsner, Kugler, Pohl & Mehrholz, 2015); Sandars, Cloutman & Woollams, 2015); Shah-Basak, Norise, Garcia, Torres, Faseyitan & Hamilton, 2015). The same facilitation may also occur with single application in healthy participants, but this remains to be established.

4. Finally, language production, and picture naming in particular, may be a good task to assess the interplay between the neurophysiological effects of tDCS and levels of cortical excitability.

The poor reliability of tDCS may be explained by differences in stimulation parameters across studies (for further discussion, see Antal et al., 2015; Horvath et al., 2016; Nitsche et al., 2015), but also by differences in baseline levels of cortical activity (see Miniussi et al., 2013). This is in part demonstrated by the generally positive effects of tDCS in brain-damaged patients, such as patients with aphasia, which contrasts with the unreliable effects in healthy samples. Following brain damage, levels of cortical excitability may become excessively low or dysfunctional compared to the optimal levels seen in healthy brains, and tDCS may help to change activation towards more optimal levels (for a similar argument, see de Aguiar et al., 2015; Miniussi et al., 2013). Furthermore, several studies with healthy participants have shown that higher baseline levels of cortical excitability can abolish the beneficial effect of anodal tDCS on task performance (see Hsu et al., 2014; Tseng et al., 2012; Berryhill et al., 2014).

Picture naming may be a good task to examine the effects of tDCS in conditions with high levels cortical excitability because this may be approximated in conditions of high semantic interference. Picture naming necessitates cortical excitation – for word retrieval – but also inhibition – for fending off competition from alternatives (for similar argument, see Miniussi et al., 2013). Moreover, the relative need of activation and selection can be manipulated by repeated presentation of semantically related pictures, which raises the general level of

activation in the lexical system while, at the same time, increasing the demand on selection. Different paradigms have been used to increase semantic interference effects, such as asking participants to name a picture when a semantically related word is present (as in picture-word interference; Mahon et al., 2007), repeatedly name sets of semantically related pictures versus unrelated pictures (as in cyclic blocked naming; for review, see Belke & Stielow, 2013), or name sets of semantically related pictures intermixed with filler items (as in continuous naming; Howard et al., 2006; Belke, 2013). In this last paradigm, for example, performance deteriorates progressively with each position of a picture in the sequence, showing the negative effects of semantic interference. Several studies have examined whether tDCS modulates these semantic interference effects (see Henseler et al., 2014; Meinzer et al., 2016; Pisoni et al., 2012; Westwood et al., 2017; Wirth et al., 2011). It has been suggested that while the excitatory effects of anodal tDCS may be generally facilitatory when applied to left frontal regions because selection abilities are boosted, when applied to temporal stimulation it may further boost activation of semantically related competitors, thereby increasing interference effects (see Pisoni et al., 2012; see also Canini et al., 2016). Finding whether tDCS modulates semantic interference will indicate whether tDCS interacts with task-induced cortical activation as well as provide evidence on the nature of interference effects.

Our review will attempt to answer the following questions:

1. *Is there a general effect of anodal tDCS targeting key language areas in the left hemisphere?* Most studies investigating language production effects apply anodal tDCS to the left frontal or temporal regions, a protocol which is assumed to give the best chances to elicit a positive effect (see Jacobson et al., 2012, for example). Therefore, we will refer to this stimulation protocol as '**conventional'** and ex- amine its effect in our ***Primary Analysis***.

2. *Is the size of the tDCS effect influenced by certain parameters?* Our ***Moderator Analysis***, therefore, assessed the impact of tDCS parameters including Timing (i.e., if tDCS was applied before or during task performance), Current Density (e.g., high vs low; .28 vs. ≥

0.057 cm/mA2), and Duration (e.g., short vs long; < 15 vs. ≥ 20 min) within conventional stimulation protocols.

3. *Is there an effect of tDCS in protocols which are not typical of the field?* Our ***Secondary Analysis*** considered the effectiveness of cathodal tDCS applied to either hemisphere and anodal tDCS applied to the right hemisphere.

4. *Is there an effect of tDCS in conditions of increased task difficulty?* Finally, the effects of anodal tDCS may be particularly evident in conditions where naming is made more difficult by the presence of competitors (possibly with the consequence of higher cortical ex- citation). Our ***Semantic Interference Analysis***, therefore, considered the effect of anodal tDCS in tasks that induce semantic interference effects, where greater effort is needed for selection and control.

3.3 Method

3.3.1 Data sampling

3.3.1.1 Eligibility criteria

Papers were included if they: a) tested healthy adult volunteers (between 18- and 60-years of age); b) included a sham control condition; c) were published in English; d) provided details of method/protocol; e) measured picture naming or word reading reaction times and/ or accuracy (given in percentage errors; or other types of accuracy scores); and f) used conventional tDCS protocols (i.e., current ad- ministered continuously via a two electrode configuration). Since the effects of tDCS are known to accumulate with repeated applications (Alonzo, Taylor, Martin & Loo, 2012; Meinzer et al., 2014), we did not include studies that applied tDCS more than once to the same cortical site with the same stimulation polarity (e.g., anodal tDCS applied over multiple days or within an hour following the first application), unless we could extract data from just the first application. Our eligibility criteria were similar to previous reviews (e.g., Mancuso et al., 2016; Price et al., 2015), but likewise broader than those used by reviews targeting studies across more diverse cognitive domains (Horvath et al., 2015a; Jacobson et al., 2012).

3.3.1.2 Literature search

We searched Science Direct, Web of Knowledge and PubMed data- bases (from 1999 to early August 2016) using as search keywords: 'tDCS' or 'transcranial direct current stimulation' in combination with 'language', 'verbal', 'linguistic', 'word production', 'naming', 'reading', and 'cognition'. We searched for further articles using the Web of Knowledge citation tracking tool, which displays articles referenced within a given article and articles that cite the article of interest. The initial search returned 3254 articles of which 2635 were removed right away as non-relevant. The text of the remaining 619 papers was read, including papers testing neurologically impaired individuals in case healthy controls were tested. This excluded 598 studies because naming or reading abilities were not measured in healthy participants, leaving 22 articles. Of these, 3 studies targeting reading were further removed because: a) recruited children and adolescents (Costanzo et al., 2016), b) did not include a sham group (Thomson, Doruk, Mascio, Fregni & Cerruti, 2015), and c) applied tDCS repeatedly but did not report data from the first application (Heth & Lavidor, 2015). Five studies targeting picture naming were also removed because: a) recruited participants were older than 60-years of age (Ross, McCoy, Coslett, Olson & Wolk, 2011); Rosso et al., 2013; Holland et al., 2011, Holland, Leff, Penny, Rothwell & Crinion, 2016), and b) collapsed data across two conditions in which tDCS targeted separate cortical regions (Manenti, Brambilla, Petesi, Ferrari & Cotelli, 2013). This left us with a final sample of 14 papers, some of which reported multiple tDCS conditions (total n = 96; within-participants, n = 86, between-participants, n = 10; see Appendix 3.1 and 3.2 for details on included studies).

3.3.1.3 Data extraction

We extracted means and standard deviations for reaction times and accuracy rates (percentage errors or other accuracy scores) for all tDCS and sham conditions reported. A java program called Plot Digitizer (Joseph, 2011) was used to convert plotted values into a numerical form if numerical values were not reported by a study (for applications of this method, see Hill et al., 2016; Vaseghi, Zoghi & Jaberzadeh, 2015). If no data was reported or could not be extracted, the authors were contacted.

3.3.2 Data analysis

3.3.2.1 Direction of the tDCS effect

As with other reviews, we quantified an effect of tDCS based on the difference in performance between tDCS and sham conditions using a standardized measure of effect size: a difference between tDCS and sham conditions divided by a measure of variability to standardize the effect (see later for details). In line with the majority of tDCS studies measuring effects on cognition, our general hypothesis was that anodal tDCS of the left hemisphere would enhance whilst cathodal tDCS of the left hemisphere would impair performance. When determining the direction of the effect, we reported effects as positive if consistent with these predictions, and negative otherwise.

Note that our *Primary analysis* still included studies which looked at semantic interference in picture naming. Here, however, we did not consider effects of tDCS on interference, but on picture naming in general (i.e., across conditions; effects on interference have been looked at separately in our *Semantic Interference analysis*). Some studies predicted a paradoxical inhibitory effect for anodal tDCS when applied to the temporal lobes – an area involved in lexical activation – in conditions of high interference. The rationale being that, in conditions with semantic distractors, anodal tDCS would boost the activation of competing alternatives as well as the target, thus making selection more difficult (see Henseler et al., 2014; Pisoni et al., 2012). This prediction, however, has been confirmed only by one study (Pisoni et al., 2012), with other studies reporting opposite outcomes (Meinzer et al., 2016) or no effect at all (Henseler et al., 2014; Westwood et al., 2017).

Thus, when we included studies measuring interference in our *Primary analyses*, we coded effects in line with our general prediction that anodal tDCS should improve whilst cathodal tDCS should impair performance. In our *Semantic Interference analysis* we looked separately at the effects for temporal and frontal tDCS. Anticipating our results, we did not detect differences by site of stimulation, thus supporting our choice of coding.

Our *Secondary analysis* wanted to investigate other less commonly used tDCS protocols. Here, we included studies that applied anodal tDCS to the right hemisphere (see Jeon & Han, 2012; Ross et al., 2010; Younger, Wagner & Booth, 2016). Given the inhibitory

relationship between left and right hemispheres (Chiarello & Maxfield, 1996; Thiel et al., 2006a, 2006b; Vines et al., 2008), we expected that, compared to sham, right-hemisphere anodal tDCS would inhibit (and thus impair) language capacities located in the left hemisphere (for a similar prediction, see Hamilton, Chrysikou & Coslett, 2011); Hartwigsen et al., 2010). Thus, we coded as positive results consistent with this outcome; negative otherwise. Other included studies applied tDCS of either polarity to the cerebellum (see Boehringer, Macher, Dukart, Villringer & Pleger, 2013; Pope & Miall, 2012). Because cerebellar nuclei are thought to inhibit frontal regions, some authors argue that the excitatory effects of anodal tDCS would impair performance on frontally mediated tasks, and vice versa for cathodal tDCS (for a similar prediction, see review by van Dun, Bodranghien, Mariën & Manto, 2016; and a study included in our review, Pope & Miall, 2012)[7]. Thus, we coded as positive results consistent with the paradoxical anodal impair versus cathodal improve outcome for studies targeting the cerebellum.

3.3.2.2 Effects from within- and between-participant studies

Most studies assessing effects of tDCS on cognition used a within- study design, where the same participants were administered sham and real tDCS. Despite this, most previous reviews calculated effect sizes as being drawn from a between-participants design (see Horvath et al., 2015a; Jacobson et al., 2012; Price et al., 2015; but also see Mancuso et al., 2016). This method, however, overestimates variance (which is reduced in a within-participants design) and, therefore, reduces the chances of finding significant results. Our review included both within- participant and between-participant studies. To increase precision, we used different methods to estimate effects for within- and between- participant designs (see Lakens, 2013; Borenstein et al., 2009).

For *between-participant* studies, we measured effect sizes using Cohen's *d* with Hedges' *g* correction. Thus:

$$\text{Cohen's d} = \frac{M_{tDCS} - M_{sham}}{SD_{pooled}}$$

[7]Note, we note that one included study by Boehringer et al. (2013) predicted a cathodal-inhibition effect. However, because the effect was zero overall the sign for this effect has no impact on the analysis.

where M_{tDCS} is the mean from the tDCS condition, M_{sham} is the mean from the sham condition, and SD_{pooled} is the pooled standard deviation, calculated as follows:

$$SD_{pooled} = \sqrt{\frac{(n_{tDCS}-1)SD_{tDCS}+(n_{sham}-1)SD_{sham}}{(n_{tDCS}+ n_{sham})-2}}$$

(where values for n_{tDCS} and n_{sham} are the sample sizes for the tDCS and sham conditions and SD_{tDCS} and SD_{sham} are the standard deviations).

Cohen's *d*, was multiplied by the coefficient *J* to give us Hedge's *g*, which corrects for the upward bias in Cohen's *d* for samples less than 20 (Hedge's & Olkin, 1985). We calculated *J* as:

$$J = 1 - \frac{3}{4df-1}$$

(where *df* is the degrees of freedom used to calculate estimate SD_{pooled}, which for two independent samples is: $df \text{ for } SD_{pooled} = n_{tDCS} + n_{sham} - 2$; where n_{tDCS} is the number of participants in the tDCS condition, and n_{sham}, the number of participants in the sham condition; see Borenstein et al., 2009). Hedges' *g* can be interpreted in the same way as Cohen's *d* – i.e., effect sizes of 0.2, 0.5, and 0.8 roughly equate to small, medium and large effect sizes, respectively.

For *within-participants* studies, effect sizes were estimated as the difference between conditions multiplied by a measure of association of scores in the two conditions and then divided by the standard deviation of the difference scores. Thus

$$\text{Hedges' } g = J = \left(\frac{\left(M_{tDCS} - M_{sham} \text{ x } \left(\sqrt{2\,(1-Corr)}\right)\right)}{SD_{diff}}\right)$$

$$SD_{diff} = \sqrt{SD_{tDCS}^2 - SD_{sham}^2 - 2 \text{ x } Corr \text{ x } SD_{tDCS} \text{ x } SD_{sham}}$$

here *Corr* is the correlation between scores in tDCS and sham conditions. Since correlations were not reported by studies, we set a conservative correlation of 0.6 based on data from several of our own studies (see Westwood et al., 2017). The review by Mancuso et al. (2016) used a similar mid-range value. All effects were calculated using Comprehensive Meta-Analysis Software V3.0.

3.3.2.3 Multiple dependent effects

In a meta-analysis, effect size estimates should be drawn from different participant samples. Violating this assumption of independence leads to an underestimation of variance and an overestimation of statistical significance (i.e., a Type 1 Error or False Positive; see Lipsey & Wilson, 2001). Previous studies have not always preserved this assumption (see Horvath et al., 2015a; Dedoncker et al., 2016a). Those reviews that preserved it selected only one effect per study, thus reducing power (see Jacobson et al., 2012; Price et al., 2015). We used composite effects for conditions carried out by the same participants where we expected similar effects of tDCS (e.g., naming/reading of different types of stimuli, such as nouns vs verbs or people vs places; online vs offline stimulation at different time intervals). We report separate effect sizes for conditions where different effects of tDCS were clearly expected (e.g., anodal vs. cathodal tDCS) and when participants carried out two tasks (e.g., reading and naming). The effect of different parameters was assessed in the Moderator analyses.

Composite effect sizes can be calculated using mean performance and variance. However, this does not consider the inter-correlation between conditions, and therefore overestimates the error term (Borenstein et al., 2009). We calculated the variance based on a formula devised by Borenstein et al. (2009; M. Borenstein, June 10, 2017 by personal communication), which accounts for inter-correlation. For example, to calculate the mean effect size and composite variance for two dependent effect sizes:

Mean $= \frac{1}{2}\ (Y_1 + Y_2)$;

Composite Variance $= \left(\frac{1}{2}\right)^2 \left(V_1 + V_2 + 2r\sqrt{V_1}\sqrt{V_2}\right)$

where V_1 and V_2 are the variances for the condition means Y_1 and Y_2, and r is the correlation coefficient – i.e., an estimate of the extent to which variances co-vary. Since r is generally not reported we assumed a plausible correlation of 0.6, based in part on our own data (Westwood et al., 2017) but also advice by Borenstein et al. (2009). Assuming a correlation of 0 means that each outcome contributes new, unrelated information to the summary effect size, thus the composite variance of two unrelated samples is half of the mean variance. This may under- estimate true variance and lead to a Type I Error (False Positive). Assuming instead a correlation of 1.0 means outcomes in one sample duplicate those in the other, thus the composite variance is just the mean variance of the two samples. This may over estimate variance and lead to a Type II Error (False Negative; see Appendix B for breakdown of composite effects)[8].

3.3.2.4 Heterogeneity

Heterogeneity refers to variation in effect sizes across studies. Such variation may arise from random sampling error, or from true differences between studies due, for instance, to variation in stimulation parameters, language domain, or target site (i.e., true heterogeneity). True heterogeneity is assumed if effect estimates differ more than would be expected from sampling error alone. True heterogeneity can question the reliability of the summary effect. The conventional test for heterogeneity is the Cochran's Q statistic. A significant Q indicates that studies differ in their estimates of effects, but it is more difficult to conclude that studies are alike from a non-significant Q because Q suffers from low power with small sample sizes. To counter this, we increased the p-value to 0.10 to exclude heterogeneity (Higgins et al., 2003). We also quantified heterogeneity as a percentage using the I^2 index.

$$I^2 \text{ index} = 100 \times \frac{Q-(k-1)}{Q}$$

[8]Comprehensive Meta-Analysis V3.0 cannot alter the correlation value – fixing it at either 1 or 0 – so the mean and corrected variance was first calculated in Microsoft Excel and then these values were imputed into CMA. Formulae for calculating the effect size presented in Effects from within- and between-participant studies still apply here.

An I2 index of 0% means variation in effect sizes is all due to sampling error, whilst an index of 100% means all variation is due to true heterogeneity. Using a rule of thumb, I2 indexes = 75%, 50%, and 25% reflect respectively high, medium and low true heterogeneity (Lipsey & Wilson, 2001).

3.3.2.5 Fixed effect vs. random effects

A fixed effect model assumes there is a true effect that is the same across all studies and that variation in the size of this effect results from sampling error alone. This assumes no heterogeneity. More weight is assigned to larger studies and less weight to smaller studies as a result. A random effects model assumes the variation across studies is also due to differences in the chosen experimental methodology, such as stimulation montage, current intensity, stimulation duration, participant design, and outcome measure (Brunoni & Vanderhasselt, 2014; Dedoncker et al., 2016a; Hill et al., 2016; Price et al., 2015). This is a more conservative assumption. We have therefore used a random effects model in our meta-analyses. With this model, every study contributes to the effect size estimate, and small studies are not given a smaller weight.

3.3.2.6 Outliers

We planned to exclude effect size estimates from each analysis which were 3 standard deviations above or below the summary effect size to avoid extreme values biasing the outcome. In all analyses, no study met this criterion, so no study was excluded.

3.3.2.7 Publication bias

Publication bias refers to the tendency to publish studies with significant results and leave in the file-drawer studies with null results. The presence of bias would question the validity of a significant effect in our review. We therefore used funnel plots to identify publication bias. These are scatter plots where effect sizes are plotted against a measure of a study's precision, such as the number of participants or, in our case, standard error. Effects from smaller (or less precise) studies should spread more around the mean effect size, while effects from larger (or more precise) studies should cluster more around the mean. In the absence of bias, the distribution will be due to sampling error alone and be symmetrical around the true effect (reflected by the mean), with the distribution of scores being smaller for

larger/more precise studies and increasingly greater for smaller/less precise studies. This will give the plots a characteristic inverted funnel shape. In the case of publication bias, instead, the distribution will be asymmetrical. Studies with fewer participants are more likely to obtain positive or negative results by chance, but, in the presence of bias, positive results will be published while negative results will be missing. We used the trim-and-fill procedure, which corrects for bias by trimming outliers and imputing effects to generate a (simulated) symmetrical distribution, thereby providing an unbiased summary effect size estimate (Lipsey & Wilson, 2001). We established the significance of bias using a method proposed by Egger, Smith, Schneider and Minder (1997)9. When we look at publication bias, we use Cohen's d effect size estimates because Hedges' g already slightly corrects for bias.

3.4 Results

3.4.1 Primary analyses

Our focus here was to assess the efficacy of what are arguably conventional protocols for targeting language areas. These include anodal tDCS applied to the left frontal or temporal regions (see Appendix 3.2, Tables 1 and 2 for a listing of conditions and how they were aggregated). Figure 3.1 and 3.2 show forest plots and summary statistics of effects on speed and accuracy scores. Effects are reported as positive if consistent with the general hypothesis that left anodal tDCS improves performance; negative otherwise. Figures report summary effects separately for reading and naming tasks and cumulating across tasks, where composite effect size estimates were used to preserve the assumption of independence in cases where participants carried out both tasks. Funnel plots following trim-and-fill correction for bias are found in Figure 3.3. Summary statistics are also provided reporting effect estimates before and after trim-and-fill along with the Egger's test for significance of publication bias.

[9]This calculates bias using the effect size estimate and the inverse of standard error (or 'precision'). A linear regression is performed on the standard normal deviate (i.e., effect size over the inverse of standard error), with the inverse of standard error serving as a predictor variable. Bias is calculated in terms of the extent to which the intercept deviates from zero (Egger et al., 1997). A significant outcome indicates bias.

There are no significant effects of tDCS on either reaction times or accuracy with the overall effects being close to 0. This is true when results are pooled across tasks and when they are considered separately. In spite of no significant results overall (even before correcting for publication bias), there is still some evidence of bias. Across both analyses the trim-and-fill procedure weakened the summary effect size. Although this effect was not significant with reaction times, it approached significance with accuracy scores.

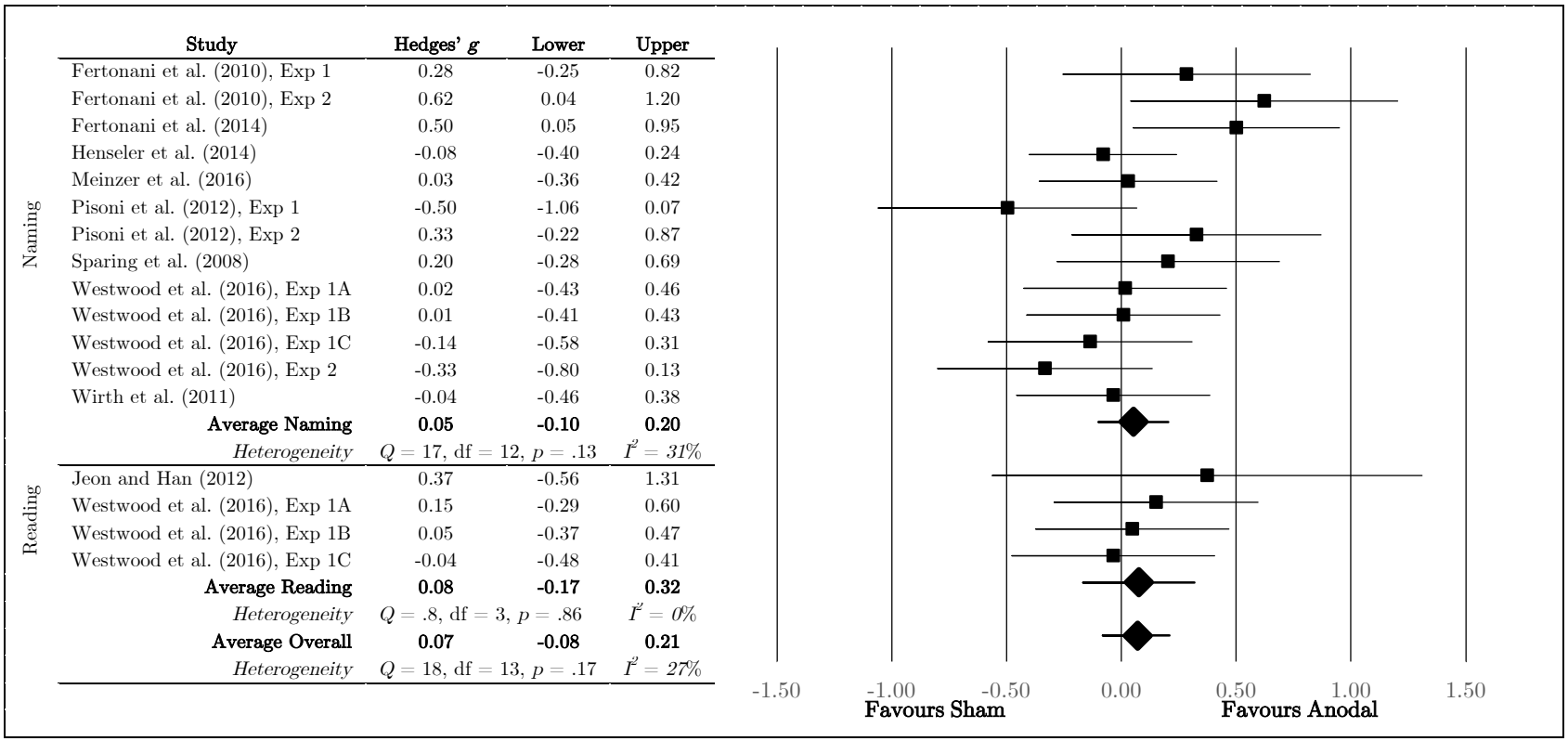


Fig 3.1 – Forest plots for the size of tDCS effects on reaction times in naming, reading and overall. Error bars reflect 95% confidence intervals. Effects size given in Hedges' g.

PRIMARY ANALYSIS: Left Hemisphere Anodal tDCS – Accuracy

Study	Hedges' g	Lower	Upper
Fertonani et al. (2014)	0.09	-0.33	0.52
Henseler et al. (2014)	-0.24	-0.56	0.09
Jeon and Han (2012)	0.59	-0.35	1.54
Ross et al. (2010)	0.00	-0.48	0.48
Turkeltaub et al. (2012)	0.18	-0.20	0.57
Westwood et al. (2016), Exp 1A	-0.13	-0.57	0.31
Westwood et al. (2016), Exp 1B	0.16	-0.26	0.59
Westwood et al. (2016), Exp 1C	-0.19	-0.64	0.25
Average Overall	-0.01	-0.16	0.14
Heterogeneity $Q = 6$, df $= 7$, $p = .52$ $I^2 = 0\%$			

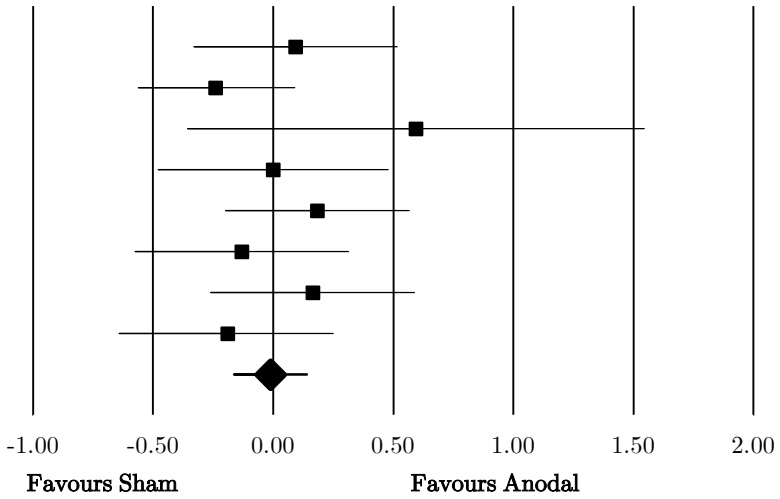
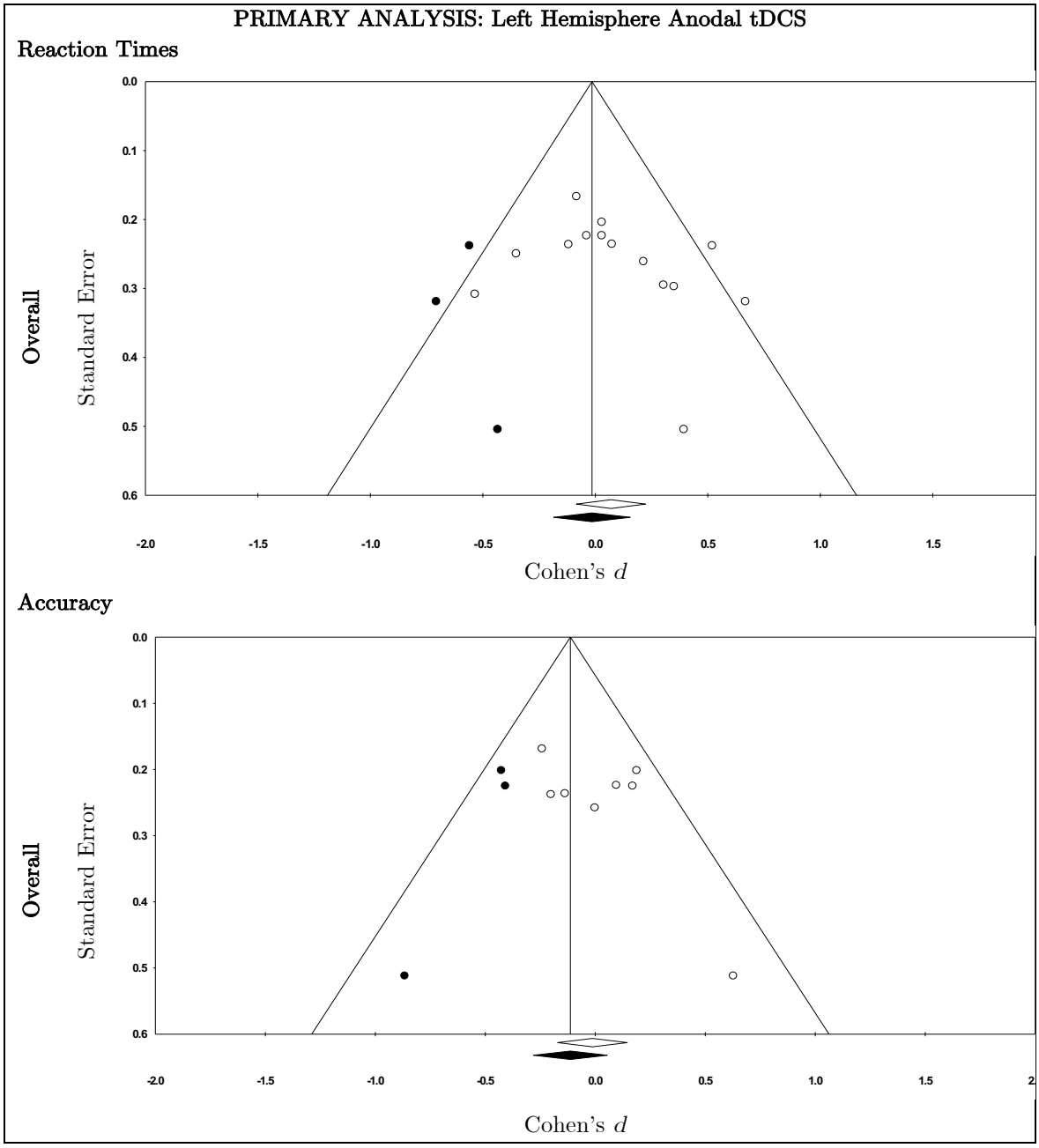


Fig 3.2 – Forest plot for the size of the tDCS effect on accuracy. Error bars reflect 95% confidence intervals. Effects size given in Hedges' g.



Summary Statistics: Publication Bias													
	Before Trim-and-Fill			After Trim-and-Fill				Egger's test					
		95% CI				95% CI			95% CI				
	d	Lower	Upper	*N*	*d*	Upper	Lower	B0	Lower	Upper	*t*	df	*p*
RTs	.07	-.08	.22	3	.02	-.14	.10	1.7	-1.33	4.6	1.2	12	.13
Acc	-.01	-.17	.14	***3***	-.11	-.28	.20	2.1	-.89	5.1	1.7	6	.06

Fig 3.3 – Funnel plots for effect of anodal tDCS using conventional parameters. Effects size given in Cohen's d. Summary statistics given in table below which shows before and after trim-and-fill effect sizes and Egger's test of publication bias.

3.4.2 Moderator analyses

Moderator analyses were carried out to identify parameters which may modulate the effectiveness of tDCS. We limited these analyses to studies which used conventional stimulation, which are more numerous. We carried out General Linear Model univariate linear regressions with size of tDCS effect (in Hedges' g) as the dependent variable and either *Timing* (*Online* vs. *Offline*), *Current Density* (0.28 vs. ≥ 0.057 cm/mA2), or *Stimulation Duration* (< 15 vs. ≥ 20 min), as the independent variables, all of which were dummy coded as categorical variables. Results are shown in Table 3.1. We found that *Timing* and *Duration* significantly moderated the tDCS effect size. Specifically, greater – yet still small – effects were observed for shorter (< 15 min) versus longer (≥20 mins) stimulation durations in terms of reaction times (M ± SE; .29 ± .08 vs. – 0.047 ± .08) and for offline tDCS versus online tDCS in terms of accuracy (M ± SE; .29 ± .11 vs. – 0.07 ± .08). It is difficult to know how much weight we should put on these significant results given the null results we report in the previous and subsequent analyses and that the effects are not observed across both reaction times and accuracy scores. Moreover, the impact of *Timing* is confounded with *Duration* as shorter durations were overrepresented in studies using offline stimulation, and vice versa. In reaction times, shorter durations make up 6 of the 10 effects for offline stimulation, whilst longer durations make up 6 of the 9 effects for online stimulation (see Appendix 3.2, Table 1); in accuracy scores, shorter durations make up 1 of the 3 effects for offline stimulation, whilst longer durations make up 3 of the 6 effects for online stimulation (see Appendix 3.2, Table 2).

Table 3.1 – Results of linear regressions for effect size estimates separately for reaction times (left) and accuracy scores (right). Significant results are highlighted in bold.

	RTs				Accuracy			
Moderator	Beta	t	p	R^2	Beta	t	p	R^2
Offline vs. Online	-0.4	-1.7	0.12	0.14	-0.72	-2.7	0.03	0.5
.28 vs. ≥.057 cm/mA^2	-0.1	-0.43	0.67	0.01	-0.55	-1.7	0.13	0.3
<15 vs. ≥20mins	-0.58	-2.9	0.01	0.33	0.22	0.6	0.57	0.05

3.4.3 Semantic interference analysis

For studies using picture-word interference (Henseler et al., 2014) and cyclic blocked naming (Meinzer et al., 2014; Pisoni et al., 2012) we calculated semantic interference as the difference in reaction times between semantically related and unrelated distractor conditions. For studies using continuous picture naming, we calculated semantic interference as the difference in reaction times between items in positions 1 and 2 and items in positions 4 and 5 in a sequence of semantically related pictures (for same method, see Westwood et al., 2017). All studies used anodal tDCS and measured reaction times. Only two studies (Henseler et al., 2014; Westwood et al., 2017) also measured se- mantic interference in terms of percentage errors, so we focused on reaction times. Because of different predictions in the case of temporal and frontal stimulation (for discussion, see Westwood et al., 2017), we also carried out separate analyses for these two conditions. Figure 3.4 shows summary effects separately for temporal and frontal stimulation and cumulating across stimulation sites, where composite scores were used to preserve the assumption of independence in cases where temporal and frontal stimulation were administered to the same participant sample (see Appendix 3.2, Table 3 for a listing of conditions and how they were aggregated). We found no effect of tDCS either overall or when considering each stimulation site separately.

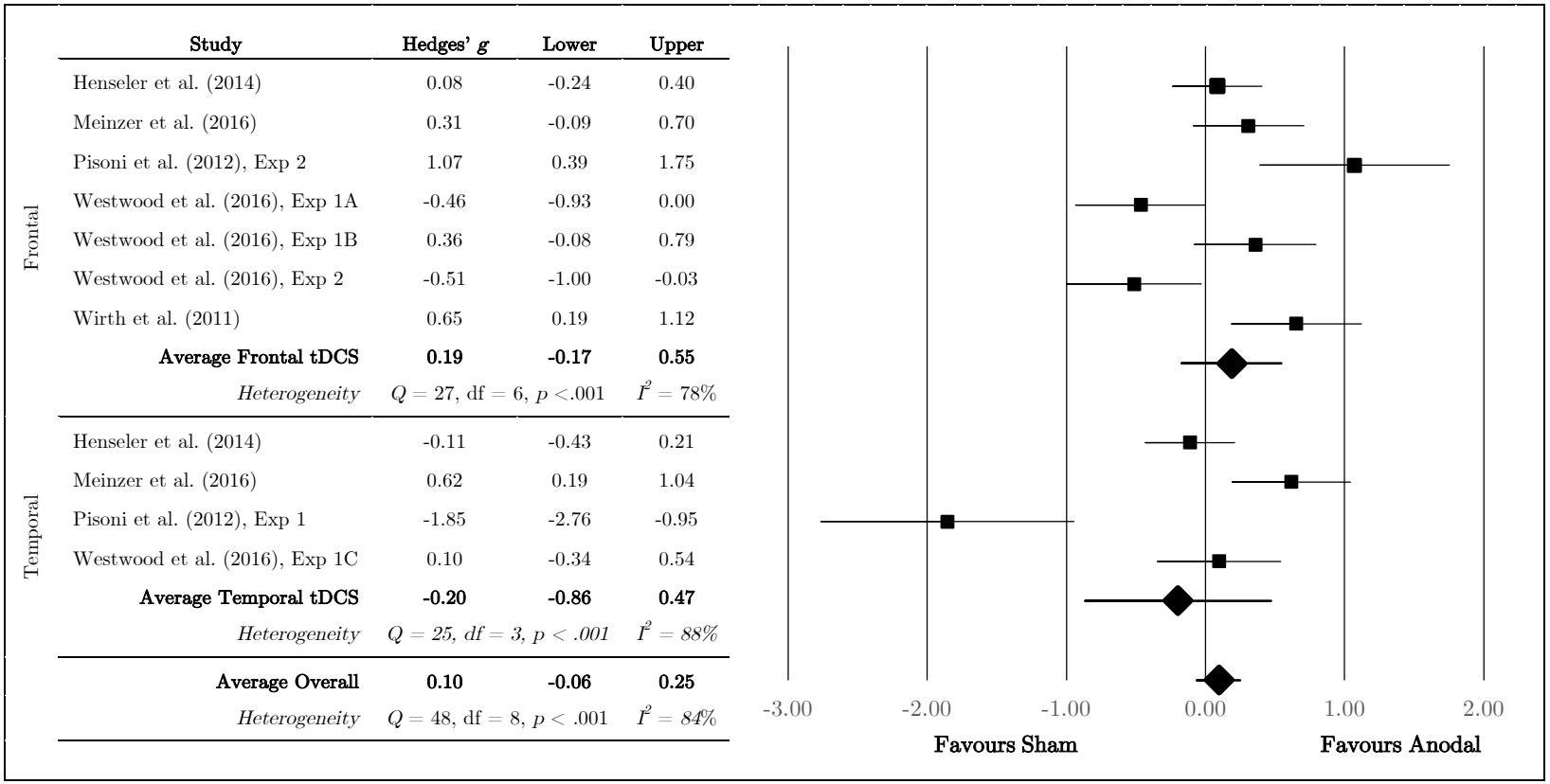


Fig 3.4 – Forest plots for the size of the effect of tDCS when considering studies measuring semantic interference effects on picture naming. Error bars reflect 95% confidence intervals. Effects size given in Hedges' g.

3.4.4 Secondary analysis

Here we explored studies that used less common combinations of stimulation polarity and locus of stimulation. These included cathodal tDCS of either hemisphere or right hemisphere with anodal tDCS. We assumed that right-hemisphere anodal tDCS would impair language capacities given the widely held assumption that right hemisphere ex- citation leads to left hemisphere inhibition (Chiarello & Maxfield, 1996; Thiel et al., 2006a, 2006b). We assumed that left-hemisphere cathodal tDCS would impair performance given its inhibitory effect on cortical excitability (see Hamilton et al., 2011; Hartwigsen et al., 2010; Woods et al., 2016), but we expect that cathodal tDCS would be paradoxically facilitatory in the case of cerebellum stimulation because cerebellar nuclei are hypothesized to exert inhibitory effects on the frontal lobes (see van Dun et al., 2016). Results are reported in Figure 3.5 (see Appendix 3.2, Table 4 for a listing of conditions and how they were aggregated). Again, we found no significant effect of tDCS for any combination of polarity and stimulation site, except for right anodal tDCS for reaction times. This effect was not expected, and should be treated with caution given the small sample size (N = 3) and that all three included studies originally reported a null effect of anodal tDCS (Jeon & Han, 2012; Sparing, Dafotakis, Meister, Thirugnanasambandam & Fink, 2008). We also estimated a significant effect in Pope and Miall (2012), contrary to the authors, who reported a null effect. This discrepancy is due to the measure we chose to estimate the tDCS effect. Pope and Miall (2012) measured reading across six trials, which were composed of five repetitions of the same stimuli followed by a sixth trial with new stimuli. The authors included all trials in their tDCS analysis, which could have diluted the effect of anodal tDCS, due to repetitions. To avoid effects of repetition, we, instead, only used performance on the sixth trial (but we could have equally chosen the first trial).

Measure	Protocol	Study	Hedges' g	Lower	Upper
Reaction Times					
	Left Cathodal tDCS	Fertonani et al. (2010), Exp 1	0.18	-0.35	0.72
		Fertonani et al. (2010), Exp 2	-0.28	-0.82	0.26
		Sparing et al. (2008)	-0.17	-0.65	0.31
		Average	**-0.09**	**-0.39**	**0.21**
		Heterogeneity	$Q = 2$, df $= 2$, $p = .47$	$I^2 = 0\%$	
	Right Anodal tDCS	Jeon and Han (2012)	-0.02	-0.95	0.90
		Pope and Miall (2012)*	-0.68	-1.28	-0.09
		Sparing et al. (2008)	-0.29	-0.78	0.20
		Average	**-0.39**	**-0.74**	**-0.04**
		Heterogeneity	$Q = 1.7$, df $= 2$, $p = .4$	$I^2 = 0\%$	
	Right Cathodal tDCS	Boehringer et al. (2012)	0.00	-0.31	0.31
		Pope and Miall (2012)*	1.37	0.72	2.01
		Average	**0.03**	**-0.24**	**0.30**
		Heterogeneity	$Q = 14$, df $= 1$, $p < .001$	$I^2 = 93\%$	
Accuracy					
	Left Anodal tDCS	Younger et al. (2016)	0.60	-0.18	1.39
	Right Anodal tDCS	Jeon and Han (2012)	-0.02	-0.95	0.90
		Ross et al. (2010)	0.00	-0.48	0.48
		Younger et al. (2016)	0.06	-0.71	0.82
		Average	**0.01**	**-0.40**	**0.34**
		Heterogeneity	$Q = .32$, df $= 2$, $p = .85$	$I^2 = 0\%$	

Fig 3.5 – Summary of effect sizes of tDCS when considering studies using atypical stimulation parameters. Lower and Upper reflect 95% confidence intervals.

3.5 Discussion

We carried out a number of meta-analyses to quantify the effects of tDCS on language tasks whilst accounting for factors that could moderate the outcome. We found no significant effect of tDCS when applied using conventional, best-evidence parameters – i.e., anodal left-hemisphere tDCS applied to frontal and temporal regions. This was true across tasks (naming and reading) and outcome measures (reaction times and accuracy). We also found no significant effect of tDCS in modulating effects of semantic interference disregarding site of stimulation (frontal or temporal), and no effects of tDCS with less used stimulation parameters – i.e., cathodal tDCS of either hemisphere or anodal tDCS of the right hemisphere. In our moderator analyses, we did find that tDCS administered offline and for a shorter duration ($<$ 15 min) produced greater effects. These effects, however, should be interpreted with caution. Effects were small (roughly .3 Hedges' *g*), confounded with one another, and contrary to predictions that greater effects should occur with longer durations (see Hill et al., 2016; see Fertonani & Miniussi, 2016; Woods et al., 2016). We believe, therefore, that more weight should be given to the large number of null findings we report, which are, overall, consistent with mounting scepticism about whether tDCS can reliably modulate cognition in healthy participants, at least with single applications (see Horvath et al., 2015a; Mancuso et al., 2016).

Given our negative results, one may ask the question: why are there so many reports of significant results across the wider tDCS literature, but also specifically across language studies? A number of factors may contribute. First of all, although we did not find any significant effect of tDCS, we still found some evidence of publication bias in our primary analysis where there is more consensus that stimulation parameters may be effective and therefore a stronger expectation of significant results. A similar bias has been found in another meta-analysis quantifying effects of tDCS on working memory tasks (see Mancuso et al., 2016). Publication bias may produce the false impression of solid effects of tDCS, even in the case of single session application in healthy participants. Secondly, reports of significant effects using conventional parameters cluster in studies where a number of conditions are run with the same participants. Since whatever effect is responsible for the better performance with real tDCS versus sham is likely to affect all conditions (whether

this is really due to tDCS or to chance factors), this will unduly inflate the significance of tDCS. This problem is well demonstrated by inspection of effects for individual conditions presented in Table B.1. We see that 6 out of 40 (or 15%) of conditions showed a significant result, but when conditions using the same participants are collapsed this drops to 2 out of 25 (or 8%). This is consistent with results being significant by chance (see also, Medina & Cason, 2017).

Finally, individual reports of positive effects of tDCS may not reflect a true effect. tDCS studies generally recruit between 20 and 30 participants, and according to one estimate the typical power achieved across cognitive studies is roughly 14%, and maybe less (see Medina & Cason, 2017). Low power naturally reduces the probability of finding a significant effect if one in fact truly exists, but it also gives undue weight to some large effects which could be significant by chance (Button et al., 2013; Minarik et al., 2016). A meta-analysis is of course an ideal tool to assess effects in fields where individual studies are underpowered. Even if we assume a small effect size of 0.25, our smaller cumulated sample (n = 160; primary analysis for accuracy) gave us good power (with a probability = 0.79) to find a significant effect if one was present, assuming a within-participants design (typical of tDCS studies measuring effects on language).

We acknowledge that the negative outcomes of our meta-analyses may be due to variability in the parameters used by different studies as well as by individual variability in the response to tDCS. It is commonly assumed that anodal and cathodal tDCS respectively up- and down- regulate cortical excitability, and that increasing stimulation duration/ density will increase the effect (see Hill et al., 2016; Kim et al., 2014; Teo et al., 2011). However, a non-linear system like the brain is unlikely to have a linear response to an externally applied electric current. First, effects may reverse with higher intensities of stimulation because of a homeostatic response (for effects on motor excitability see Batsikadze et al., 2013; Fertonani and Miniussi, 2016). Additionally, the effect of the current may interact with the present level of cortical excitability which may, in turn, depend on task demands (Fertonani & Miniussi, 2016; Miniussi et al., 2013), and/or individual differences in base-line levels of excitation or cognitive ability (see, Hsu et al., 2014; Tseng et al., 2012; Berryhill et al., 2014; Bikson & Rahman, 2013; Krause et al., 2013). We considered semantic interference – a

proxy for heightened cortical excitation relative to normal cortical activity – but still found no evidence of significant tDCS modulation (see also Westwood et al., 2017). Our review, therefore, suggests that we are yet to find the conditions in which the level of cortical excitability is optimal for improving performance, at least in language tasks within a single session of tDCS and in healthy participants.

It is also important to stress that we focused on studies recruiting young healthy participants. In contrast to our null results, positive effects of tDCS in naming tasks have been consistently noted in aphasic patients (for review, see Cappon et al., 2016; Crinion, 2016; Sandars et al., 2015). Positive effects have also been reported in older adults, although less consistently (see Fertonani et al., 2014; Ross et al., 2011). It is possible, therefore, that positive effects are much easier to elicit in populations where levels of cortical excitation are suboptimal due to brain damage or aging. Finally, our investigation was limited to picture naming and word reading tasks. It is possible that single applications of tDCS cannot modify processes and/or representations involved in these tasks since they are so well established through years of practice. Positive results, instead, may be achieved in other tasks where more novel processes are engaged. Learning paradigms, for example, may provide more positive results, even in control participants (Flöel et al., 2008; Fiori et al., 2010; Meinzer et al., 2014), because here, as in the case of aphasic patients with brain-damage, representations are weaker and in a more 'plastic' state. Alternatively, cognitive effects in healthy participants may be reliable only when tDCS is administered repeatedly with cumulative effects (Alonzo et al., 2012; Meinzer et al., 2014).

3.5.1 Conclusions

Undoubtedly our results are not encouraging regarding the ability of tDCS to modulate cognitive performance in a single session with healthy participants. It is too early, however, to conclude that tDCS is generally ineffective in this population. Future studies should investigate tDCS effects on tasks which involve learning and/or involve repeated application of tDCS. Future studies should also continue to investigate interactions with underlying levels of cortical excitation. Historically, novel interventions pass through a hype cycle – i.e., an initial peak of interest which then wanes with growing scepticism – before conditions

in which the intervention can operate reliably are established. Hopefully, the same will occur with tDCS. In this endeavour, however, it is very important to have a fair assessment of the limits of this technique and of the conditions in which there is no or limited efficacy. We have already learned from another form of brain stimulation, transcranial magnetic stimulation (TMS), that the torrent of what later turned out to be false positive or unreliable reports generated so much noise that it slowed the uptake of TMS in conditions where effects are indeed reliable, such as in clinical depression (see Walsh, 2013). We hope our review will help in establishing the right scope of application of tDCS.

Chapter 4: Investigating the effects of tDCS on verbal fluency and working memory

4.1 Overview

In this chapter, I assessed whether single session of LIFG anodal tDCS could modify verbal fluency and working memory in healthy participants across two novel experiments. I measured tDCS effects across the overall task performance, and considered whether the effects of tDCS would differ according to task related demands. Models of language production see verbal fluency as being underpinned by activation and inhibition of the target and competitors respectively, with inhibition particularly helpful to facilitate switching strategies that help with selection of the lexical semantic representations. Models of working memory, on the other hand, state that because of its transient storage capacity, efficient working memory relies on the ability to inhibit interference from irrelevant information. Moreover, because the capacity to exert inhibitory control is likely to differ between participants, I considered baseline ability (e.g., working memory capacity, interference control), in addition to looking at the impact of cortical excitability, individual variation, and task demands. Despite these efforts, no effects of tDCS were observed. I submit below an corrected version which was submitted[10] to *Frontiers in Neuroscience* as part of a special issue. The formatting (e.g., headings, labels, appendices, citations and references) were changed to be consistent with the style of the thesis. I also include supplementary material that was not included in the published manuscript.

4.2 Introduction

Transcranial direct current stimulation (or tDCS) is a non-invasive form of brain stimulation which is used to modulate cognitive performance by applying a weak electric current via

[10] Westwood, S. J., & Romani, C. (2018). Null Effects on Working Memory and Verbal Fluency Tasks When Applying Anodal tDCS to the Inferior Frontal Gyrus of Healthy Participants. *Frontiers in neuroscience*, *12*, 166.

electrodes placed on the scalp. Early studies measuring effects of tDCS on motor cortical excitability suggested that the applied current can cause directional changes in the resting membrane potentials underneath the electrodes—with predominant depolarization under the anode (known as anodal tDCS) vs. hyperpolarization under the cathode (cathodal tDCS; de Berker et al., 2013). It is widely assumed that effects on cortical excitability map on to cognitive effects, with anodal vs. cathodal tDCS improving vs. worsening the cognitive function of targeted brains regions. However, though widely assumed, this might not necessarily be the case. Current flows between the electrodes with complex effects that are poorly understood. Moreover, an important confounding factor modulating the impact of tDCS may be individual variation in cortical activity and/or level of ability (for reviews, see Miniussi et al., 2013; Horvath et al., 2015; Li et al., 2015; Westwood & Romani, 2017; Westwood et al., 2017). These are widely cited as explanations for a number of recent reports of negative, inconsistent, and/or small effects linked to single applications of tDCS especially in healthy participants (see Horvath et al., 2015; Mancuso et al., 2016; Westwood et al., 2017). Our study will contribute to clarify the scope of tDCS effects by considering tasks that tax executive selection abilities, mediated by the frontal lobes, and where positive, but inconsistent, effects have been reported before. We will consider effects on the whole participant group, but crucially also on subgroups subdivided according to (a) general performance and control abilities; (b) working memory span; and (c) motivation levels to see whether these variables affect tDCS outcomes.

We will tap executive selection using verbal fluency tasks and probe tasks. In fluency tasks, participants have to name in 60 s as many unique words as possible that belong to a given semantic category (semantic fluency) or begin with a given letter (phonemic fluency; for review, see Whiteside et al., 2016). In probe tasks, participants judge yes or no whether a test item (or probe) was present in a target list presented immediately before (e.g., referred to as *Sternberg task*; *recent-probe*; *Deese-Roediger-McDermot*, or *DRM*) or in a particular position in a continuous sequence of items (e.g., the n-back task; for reviews, see Jonides and Nee, 2006; Irlbacher et al., 2014). We will attempt to modulate executive selection on these tasks by targeting the left inferior frontal gyrus (LIFG) with anodal tDCS. Various lines of research suggest that this brain region plays an important role in supporting performance on these tasks, and in executive selection processes more generally

(see Hirshorn and Thompson-Schill, 2006; Jonides and Nee, 2006; Badre, 2008; Nelson et al., 2009; Atkins et al., 2011; Robinson et al., 2012; Biesbroek et al., 2016), with positive effects reported with tDCS and other forms of non-invasive brain stimulation (Feredoes et al., 2006; Price et al., 2015; Hill et al., 2016).

One expectation is that we will find a beneficial effect of anodal tDCS on task performance, but that possibly this effect will not be uniform across participants. Optimal executive selection is a dynamic interplay between automatic activation and controlled modulation of this activation—e.g., some activated responses will be selected whilst others are suppressed in the service of a goal (Thompson-Schill et al., 2005; Barak and Tsodyks, 2014; Sprekeler, 2017). Individual differences in the capacity to recruit control mechanisms will potentially interact with the tDCS effect resulting in either a net positive or negative outcome (Krause et al., 2013; Krause and Kadosh, 2014). For example, one possibility is that if selection is operating at optimum levels at baseline, anodal tDCS may have no effect or may increase excitability beyond the optimum working range, but, if selection is poor at baseline then tDCS may boost this ability (as well as increase general activation levels) with net positive outcomes. This is consistent with evidence that effects of anodal tDCS are determined by working memory span (see Berryhill and Jones, 2012; Jones and Berryhill, 2012; Berryhill et al., 2014; Jones et al., 2015; Gözenman and Berryhill, 2016), and baseline levels of inhibitory control and task ability (Sela et al., 2012; Tseng et al., 2012; Hsu et al., 2014, 2016; Jones et al., 2015; London and Slagter, 2015). The impact of such individual differences will additionally be compounded by task mediated demands on executive selection processes.

We chose to target fluency and probe tasks because they are particularly apt for exploring task mediated variation in executive selection. In verbal fluency, some areas of the lexicon will be activated, but participants will have to carefully match items to selection criteria, whilst inhibiting earlier responses. More importantly, because participants prefer to produce clusters of words similar in meaning (e.g., dog, cat, mouse) and/or sound (e.g., lift, link, listen), exhausted clusters need to be inhibited whilst a new selection criterion is generated in order to switch to a new cluster (see Shao et al., 2014; Berberian et al., 2016; Whiteside et al., 2016). In probe tasks, one can devise conditions that introduce lure probes that are either related (e.g., semantically or associatively) to items in the target list (e.g.,

the *DRM* task or a variant used in this study, the *semantic-associated probe*), were presented in a previous list (e.g., *recent-probe*), or—in the case of the n-back task—one can place targets next to the target position in the sequence (e.g., placing the target 3-back in a 2-back task, such as the target j in the sequence j, a, b, j in a 2-back task; see seminal work by Gray et al., 2003). In both cases, control resources must be deployed to update contents in working memory and to suppress lures which will otherwise bias responses due to their relatedness or familiarity with list items (Jonides and Nee, 2006; Novick et al., 2010; Atkins et al., 2011; Irlbacher et al., 2014).

One key aim of this investigation is to subdivide participants on a number of measures to see whether differences on these measures can predict differences in response to tDCS. Firstly, we will use general measures of performance in terms of overall performance on verbal fluency and probe tasks and digit span. Secondly, we will use more direct measures of executive control. For the fluency tasks, we will divide participants based on the number of correct switches over the total number of correct responses at baseline; the assumption being that—in line with previous studies—greater switching reflects better control abilities (see Troyer et al., 1998; Hirshorn and Thompson-Schill, 2006). For the probe tasks, we will consider the difference in performance between lists containing neutral vs. lure probes, the assumption being that a smaller interference from lures reflects better control abilities (see Hirshorn and Thompson-Schill, 2006; Jonides and Nee, 2006; Irlbacher et al., 2014; Shao et al., 2014). Finally, we will use a measure of motivation to succeed on a task because it has been shown previously that participants who score higher in this trait perform better on working memory tasks (for review, see Fino et al., 2014) and are more amenable to tDCS modulation (see Metuki et al., 2012; Sela et al., 2012; Jones et al., 2015). For this we will use the BAS component of BIS/BAS scale (Behavioral Approach System/Behavioral Inhibition System, Carver and White, 1994), which measures the reward sensitivity trait (for similar method, see Metuki et al., 2012; Sela et al., 2012), and is correlated positively with working memory and cognitive control abilities more generally (Gray and Braver, 2002; for reviews, see Gray and Burgess, 2004; Savine et al., 2010; Fino et al., 2014).

Before moving to our experimental investigation, we will now briefly review existing studies assessing the effects of tDCS on verbal fluency and probe tasks.

For *verbal fluency*, early reports found promising evidence that applying anodal tDCS to the left prefrontal cortex for up to 20 min can increase the average number of words produced (Iyer et al., 2005; Cattaneo et al., 2011). However, not all studies reported positive results (see Cerruti and Schlaug, 2009; Vannorsdall et al., 2012, 2016; Penolazzi et al., 2013a; Binney et al., 2018). In a meta-analysis, Price et al. (2015) found small to moderate effects (roughly 0.5, Hedges' g) for anodal tDCS in studies measuring verbal fluency (n = 6) or language learning (n = 2) when pooling all studies together. Positive effects were also found for studies measuring offline effects on verbal fluency (n = 3). However, significant effects were potentially carried by three effect size estimates, which were exceptionally large relative to others (0.8 = Flöel et al., 2008; 1.1 = Cattaneo et al., 2011; 0.7 = Meinzer et al., 2012). Studies generally find more significant effects of anodal tDCS with semantic compared to phonemic fluency (see Cattaneo et al., 2011; Horvath et al., 2015; Price et al., 2015). Only three studies have measured clustering and switching. Two applied anodal tDCS to left frontal regions, with one study showing an increase in cluster sizes in semantic fluency (Vannorsdall et al., 2012), whist another showed no effect at all (Penolazzi et al., 2013b). The third study targeted dorsal-frontal, temporal-parietal, and frontal-temporal regions, and found that only cathodal tDCS applied to the frontal-temporal regions increased cluster sizes (see Binney et al., 2018). However, unlike previously mentioned studies, this last study used a three electrode montage which limits comparisons (e.g., two cathodes placed bilaterally over the left *and* right hemisphere).

For *working memory*, a number of reviews report similarly mixed results. One review reported anodal tDCS related gains on both reaction times and accuracy scores (see Hill et al., 2016), another only on reaction times (see Brunoni and Vanderhasselt, 2014). A more comprehensive review found small to null effects across reaction times and accuracy scores following their own meta-analysis and a re-analysis of the two previous meta-analyses (Mancuso et al., 2016). Instead, a significant but small effect was seen for working memory training (Mancuso et al., 2016). However, a majority of tDCS studies measuring effects on working memory do not directly measure performance during lures trials. Since lure trials place a greater load on hard to recruit control mechanisms, tDCS has more scope to modulate performance, as discussed previously. This may explain the small to null effects

reported in the above meta-analysis, which pool predominantly from studies using n-back tasks without lures, and the positive effects seen on probe tasks that include lures (such as the modified recent-probe task and semantic associated probe tasks).

Gladwin et al. (2012), for instance, reported that anodal tDCS to the *left dorsolateral prefrontal cortex* (dlPFC) decreased reaction times on lures trials in their modified version of the *recent-probe*task, but no effect was found on neutral trials. By contrast, when using the *Sternberg* task (which does not include lures), two studies reported null effects (see Mulquiney et al., 2011; Teo et al., 2011), another found that cathodal but not anodal tDCS improved performance (e.g., Ferrucci et al., 2008), whilst oscillatory anodal *and* cathodal tDCS worsened performance in another study (Marshall et al., 2005). However, effects in this last study might be attributable to unconventional stimulation parameters (e.g., bifrontal tDCS, with intermittent stimulation see also Discussion). Positive effects on lure trials but not on non-lure trials were also reported on a modified n-back task, when applying High Definition (or HD) tDCS to the dlPFC. However, given that HD-tDCS uses a multi-electrode array that improves current focality, it is difficult to infer whether this positive effect was mediated by the presence of lures or stimulation parameters (see Hussey et al., 2015).

Other studies report similar effects targeting the *temporal* or *parietal* regions, two regions that support performance on probe tasks that include semantically and/or associatively related lures (for review, see Lambon et al., 2001; Jefferies and Lambon Ralph, 2006; Binder and Desai, 2011; Jefferies, 2013; Mirman and Britt, 2014). One study reported that anodal tDCS applied to left anterior temporal lobe can decrease false alarms for semantically related lures (Boggio et al., 2009), whilst another reported a decrease in false alarms for associative, but not semantically related lures (Díez et al., 2017). Other studies showed an increase in hits (i.e., correct *yes* responses to probes) when targeting the parietal cortex with a bilateral montage (i.e., right-anodal/left-cathodal), whilst the opposite montage (i.e., left-anodal/right cathodal) *increased* false alarms (see Pergolizzi and Chua, 2015, 2016; but see also, Pergolizzi and Chua, 2017).

Aims of Study

In our experimental investigation, we intend to stimulate the LIFG to evaluate factors that may drive the inconsistent effects seen in verbal fluency and probe tasks. In regards the latter, we chose to focus on recent-probe and semantic-associated probe, since studies using similar tasks have found positive effects of anodal tDCS, and because one can measure the differential impact of tDCS on control mechanisms during performance on non-lure and phonological and semantic lures in similar types of tasks. No study to our knowledge has applied anodal tDCS to the LIFG in these tasks, despite evidence from studies using other forms of non-invasive brain stimulation (Feredoes et al., 2006), and the role this region plays in switching and interference resolution on lure trials, and in executive selection more generally (for reviews, see Jonides and Nee, 2006; Badre and Wagner, 2007; Badre, 2008). We will measure performance across the whole group of participants, but also subdivide participants based on different measures which may modulate the effect of tDCS. We hypothesize that anodal tDCS would improve performance across tasks, but this may change in accordance with individual and task mediated variation in executive selection, with a preferential effect on individuals with suboptimal executive selection abilities since these are more likely to be the beneficiaries of a boost potentially provided by tDCS.

4.3 Experiments 1 & 2: tDCS effects on fluency and probe tasks

4.3.1 Method

4.3.1.1 Design

All participants completed two testing sessions 1 week apart. Experimental participants carried out one session with active tDCS and one with sham tDCS. Control participants carried out two sessions without any form of stimulation. All participants carried out two parallel versions of the semantic and phonemic fluency tasks (*Experiment 1*) and two parallel versions of a *recent-probe* and a *semantic-associated probe* task *(Experiments 2*) across the two testing sessions. For the experimental participants, one group carried out the recent-probe whilst another group carried out the semantic-associated probe, but both groups carried out the phonemic *and* semantic fluency tasks. The control participants carried out all tasks. In the first session, the experimental participants were also administered a digit span task and the BIA/BAS scale before stimulation.

The time taken to complete one version of a given probe task and the verbal fluency tasks was roughly 20 min. We counterbalanced the order of session, stimulation, and task version across participants. To avoid participants using words presented in the probe task for their responses in the fluency tasks, fluency was always performed before the probe tasks.

4.3.1.2 Transcranial direct current stimulation (tDCS)

Stimulation was administered via a battery driven NeuroConn DC-Stimulator using a 25 cm^2 anode and 35 cm^2 cathode inserted in sponges soaked in saline solution. The anode was placed on the LIFG, whilst the cathode was placed on the contralateral supraorbital area. The LIFG was located as F7 in the 10/20 EEG system, which we located by measuring 2 cm from the corner of the eye to the ear then 3 cm at perpendicular upwards (see Gough et al., 2005). We administered a 1.5 mA current for 25 min. Stimulation was administered 5 min before participants performed the first (fluency) task, and continued throughout the duration of the other tasks (for the same method, see Westwood et al., 2017). These parameters were in line with previous studies (see Appendix 4.1). To assess the integrity of blinding, participants were asked about their experience of tDCS via a feedback questionnaire at the end of each session (see Fertonani et al., 2010).

4.3.1.3 Tasks

Verbal Fluency. Parallel versions of semantic and phonemic fluency tasks were used. For phonemic fluency, we used the letters: *C, L, S, A*; for semantic fluency we used categories: *Animals*, *Fruits*, *Super Market Items*, *Musical Instruments* (for similar procedure, see Cattaneo et al., 2011). Letters were chosen from the two most widely used phonemic fluency tasks (i.e., *FAS* and *CLF*; for review, see Barry, Bates & Labouvie, 2008). Semantic categories were chosen because they were similar to those used in others tDCS experiments (see Cattaneo et al., 2011; Vannorsdal et al., 2012). Our selection of stimuli were justified in the control participant data, which showed good correspondence across versions.

Participants were given one minute to name as many unique words as possible that started with a given letter or belonged to a give semantic category. Proper names (e.g., *Rochester* or Robert) or repetitions (even with a different ending; e.g., *eat* followed by *eating*) were

not allowed. Participants were reminded to keep going until the time ran out even if they drew a blank. To ensure participants understood the task, the experimenter provided a practice (e.g., "*for the letter T I could say, 'terrible', 'turn', and 'table'*) and asked participants if they could think of any other words. After instructions were given participants were asked if they had any questions. Responses were recorded using a voice recorder and scored after the testing session. Our primary outcome measure was the average number of words produced correctly, with repetitions and rule violations excluded. Slang words and foreign words were permissible answers so long as they as they were listed as Standard English. Participants were asked to indicate the meaning of a word in instances of ambiguity (e.g., *frank*) at the end of the task (for a similar procedure, see Cattaneo et al., 2011; Iyer et al., 2005; Penolazzi et al., 2013a). We measured cluster sizes and switching using the protocol designed by Troyer, Moscovitch and Winocur (1997; see also Toyer, 2000; Troyer & Moscovitch, 2006). Briefly, for phonemic fluency, clusters were two or more words generated consecutively that either a) shared the same first two letters (e.g., clay, cliff); b) shared the first and last sound, but differed by only one vowel sound (e.g., cloak, clock); c) rhymed (e.g., chest, crest); or d) were homonyms (e.g., cache, cash). For semantic fluency, a cluster were words generated consecutively that belonged to a subcategory (e.g., *farm animals*, *string instruments*, *citric fruits*, *dairy products*; for full details, see Appendix 4.3). Switches were the number of transitions between clusters, including single words.

Probe Tasks. Participants were shown a list of words before making a yes/no decision about whether a test word, or *probe*, had appeared in the list. Responses were given by pressing keys *g* (for *yes*) or *j* (for *no*) using the index finger of the right hand. Participants were asked to give fast and accurate responses. List items were presented one after the other, each centred for 800 msec followed by a blank screen for 500 msec then another item was presented; probes were presented centred for 4000 msec or until participants gave a response after which a blank screen followed for 1500 msec before the next trial started. Words were presented in black Courier New typeface 18-font. Probes appeared in red ink to distinguish them from list items. Words were presented using E-Prime 2 Software and a Dell Laptop computer screen (screen size: 15.6"). We matched words across list positions in terms of word length, and across probe types in terms of frequency (based on CELEX Database; Baayen et al., 1995), and word length (see Appendix 4.2). Lists (including paired probe)

were presented randomized for the semantic-associated probe, but were presented in the same order for the recent-probe because the list order is important to ensure lure probes appeared *n* number of lists back. In terms of scoring, for reaction time analysis, we excluded incorrect responses and reaction times below 250 msec or above 2.5 standard deviations from the participant mean.

Recent-probe stimuli. A set of 216 randomly selected words were repeated two or three times to generate a list of 408 words. From this list we generated 51 word lists each composed of 8 words plus one probe word. There were two types of probes: *positive* (appeared in the word list; $n = 21$) and *negative* probes (did not appear in the list; $n = 30$). There were three types of negative probes, *negative* (which did not appear in the preceding two lists; $n = 10$), *recent-negative* (appeared in the immediately preceding list*;* $n = 10$) and *non-recent-negative* (appeared in the previous but one list; $n = 10$). Probes were never items presented in the eighth position of a given list; positive probes were taken from each list position (3 probes per position); negative probes were randomly selected from each position (1 or 2 items per position); and positive and negative probes never appeared in the preceding or preceding but one list. For the parallel version, I resampled the stimuli to generate a new set of 51 word lists and paired probe items. We matched both versions across list positions in terms of word length, and probes in terms of frequency, (based on CELEX Database; Baayen et al., 1995), and word length (see Appendix 4.2).

Semantic probe stimuli. We generated 160 word lists each composed of 5 words plus one probe word from a pool of 982 nouns. There were *Positive* ($n = 90$) and *Negative* ($n = 90$) probes lists. Positive probes, were either *positive-related* ($n = 40$) and *positive-unrelated* ($n = 50$). For *positive-related*, the list included the probe plus one word semantically related to the probe (e.g., *plug, tunnel, wire, bishop, bracelet*; probe: *plug*); for *positive-unrelated*, the list included only the probe with no other related word (e.g., *bandage*, *shield*, *life*, *puff*, *worker*; probe: *bandage*). For negative probes, there were *negative-associated* ($n = 20$), *negative-combined* ($n = 20$), *negative-associated-combined* ($n = 20$) and *negative-unrelated* ($n = 30$). For ***negative-associated***, the list included two items semantically related to each other and to the probe (e.g., *valley, plum, violin, peach, shawl*: probe: *apricot*)*;* for ***negative-combined***, the list included two words that were unrelated to one another but whose meaning

overlapped with the probe (e.g., *vehicle, lobe, lizard, jewel, hostage*: probe: *earring*), for ***negative associated-combined*** the list included two words which were related to one another *and* whose meaning overlapped with the probe (e.g., *cage, **book**, law, **plot**, plot*: probe: *novel*). For ***negative-unrelated***, the list did not include any items semantically related or whose meaning overlapped with the probe (e.g., *ball, table, wire, camel*: probe: *boat*). The order in which words were presented in each list was the same across participants. The selection of probe items and the list position of items related to the probe were controlled in the following manner. For positive-unrelated probes, 6 positive probes were taken from each position (i.e., 6 probes x 5 positions = 30 probes); for positive-related probes, 8 related items were positioned in each of the first 4 list positions (i.e., 8 probes x 4 positions = 30 probes); for negative-associated and negative-combined probes, the two related items appeared always in the second position and then either in the second or third position an equal number of times. For parallel versions of the task, I separated the lists equally between the two versions. The two versions were matched for position list in terms of word length, and across probes in terms of frequency, (CELEX Database; Baayen et al., 1995), and word length of probes (see Appendix 4.2).

Digit span. At the beginning of the first session, experimental participants completed the digit span task. In this task, participants were given a sequence of single digits and asked to recite the sequence in the order it was given. The experimenter read each digit sequence aloud with about a 1s interval between each digit. The task started with a set of *four*-digit sequences, and if more than five of the ten sequences were recalled correctly, the participant was given ten *five*-digit sequences, and so on until the final set of *eight*-digit sequences was completed or they could not accurately recite more than five digit sequences. Each sequence recalled correctly was scored as 1 (10 correct sequences = 10 points). Span was the sum of scores obtained for each length.

Motivational Scale (BIS/BAS). After the digit span, participants completed the BIS/BAS questionnaire (Carver & White, 1994). This questionnaire is designed to measure two dimensions of trait motivation: the Behavioural Approach System (BAS), which measures responsiveness to reward, whilst the Behavioural Inhibition System (BIS), which measures responsiveness to aversive stimuli (see Carver & White, 1994). The BIS/BAS scale asks

participants to rate to what extent they agreed or disagree with twenty-four statements. Responses were given using a four-point rating scale, which ranged from 1 ('very true for me') to 4 ('very false for me'). Participants were asked to respond to all items as accurately/honestly as possible, providing only one response to each item. It was stressed that each item should be considered on its own, so as to avoid participants making their responses 'consistent'. The BIS provides one measure of behavioural inhibition (*BIS*; e.g., ''I have very few fears compared to my friends"), the BAS provides three measures of reward sensitivity: 1) reward responsiveness (*BAS-RR*; e.g., ''It would excite me to win a contest''); 2) Drive (BAS-D; e.g., ''I go out of my way to get things I want''); and, 3) Fun Seeking (FAS-FS; e.g., ''I often act on the spur of the moment''; for further details on scale item, see Appendix 4.4)

4.3.1.3 Participants

Sixty-three undergraduate students from Aston University participated for course credits or financial reimbursement, and were assigned to the experimental or control group in a semi-random fashion. One participant in the experimental group failed to turn up to the second stimulation session due to other commitments. This left twenty participants for recent probe (11 female; 20 ± 1.10); nineteen participants for the semantic associated probe (9 female; 19 ± 1.00); and twenty-four participants for the control group (9 female; 21 ± 1.20). All participants were right-handed and native English speakers. We excluded volunteers with language impairments, history of migraine, headaches (frequent or severe), skin disorders (e.g., eczema), any adverse experience to previous tDCS, any history of epilepsy or stroke, head/metal implants, any neurological disorders, and any volunteers who had participated in a tDCS or TMS study in the 6 months prior to the current study.

4.3.1.3 Data re-sampling

Pseudo stimulation conditions from control data. For the experimental participants, we counterbalanced the order of stimulation (Sham vs. tDCS) and the task stimuli sets (A vs. B). Thus, in session one, half of participants received sham whilst the other half received active tDCS, and half of participants that received either form of stimulation saw stimuli set A whilst the other half saw B. In the control group – in which stimulation was not

applied – half of participants saw set A or B in the first session (and vice versa in the second session). Thus, to make data from the control group comparable with data from the experimental group, we resampled data from the control group to create two pseudo datasets (referred to as, pseudo-sham and pseudo-real), with each dataset including data from the first and second testing session and from stimuli sets A and B (for the same method, see Westwood et al., 2017).

Division of participants into sub-groups. We generated subgroups based on the median baseline scores of working memory span (digit span scores), motivation (BAS-RR scores), switching (switches over total words generated) and interference control. Switching and interference control were based on data recorded during sham stimulation (for similar method, see Hsu et al., 2016). Interference control was based on probe task performance, and calculated as the difference between aggregated performance on negative trials versus lure trials. For recent-probe, we calculated the difference between negative probes and the average of recent-negative and non-recent-probes; for the semantic-associated probe, we calculated the difference between negative unrelated probes with the average of negative-associated, -combined and associated plus combined trials.

4.3.2 Results for Experiment 1: Verbal Fluency

4.3.2.1 Group analysis

Overall performance. Figure 4.1a shows overall performance in terms of the average correct number of responses generated for semantic and phonemic fluency tasks across stimulation conditions and participant groups. We observed that participants generally produced more responses in the semantic fluency, which was expected since phonemic fluency is comparatively more effortful (for similar results, see Cattaneo et al., 2011; Vannorsdall et al., 2016), but importantly in no instance did tDCS improve performance relative to sham.

We compared separately the overall difference between stimulation conditions for each participant group via individual one-way ANOVAs, with *Condition* (Real vs Sham for experimental group; Pseudo-Real vs Pseudo-Sham for control group) as a repeated-participants factor, and the number of correct responses generated as the dependent

measure. The results showed no significant main effect of *Condition* ($F(1,38) = .39$, $p = .54$, $\eta_p^2 = .01$; $F(1,23) = .32$, $p = .58$ $\eta_p^2 = .01$).

We combined results from all participant groups to carry out a mixed factor ANOVA, with *Condition* (Real vs Sham for experimental group; Pseudo-Real vs Pseudo-Sham for control group) and *Fluency Task* (Phonemic vs Semantic) as within-participants factors, and *Group* (Control vs Experimental) as a between-participants factor. There was a significant main effect of *Fluency Task* ($F(1,61) = 102$, $p < .001$, $\eta_p^2 = .63$), with a greater number of responses generated in semantic fluency compared to phonetic fluency, as was expected (21.3±.5 vs 27.7±.6, respectively). There was a significant main effect of *Group* ($F(1,61) = 5.83$, $p = .02$, $\eta_p^2 = .09$), with experimental participants providing more responses across both tasks than control participants (26±.56 vs 23±.72, respectively). Importantly, there was no significant interactions, including *Condition* by *Group* ($F(1,61) = .66$, $p = .42$, $\eta_p^2 = .01$), *Condition* by *Task* ($F(1,61) = .96$, $p = .33$, $\eta_p^2 = .02$), *Task* by *Group* ($F(1,61) = 2.41$, $p = .13$, $\eta_p^2 = .04$), and *Condition* by *Group* by *Task* ($F(1,61) = .35$, $p = .56$, $\eta_p^2 = .01$).

There was clearly no significant effect of stimulation. We did find that the experimental group generated more responses compared to the control group, but this difference did not interact with stimulation condition. We assume this group difference arose from a placebo benefit in the response to the *presence* – not the effect – of tDCS, which raised arousal levels.

Effect of stimulation on switching versus clustering. Here, we considered that the impact of tDCS may vary according to the average number of switches and the average size of individual clusters. Figure 4.1b and 1c shows the average number of switches and average size of clusters across fluency tasks, stimulation conditions and participant groups, and it is again clear there was not effect of tDCS on switching or cluster sizes.

We compared separately the difference between stimulation conditions for each participant group via individual one-way ANOVAs, with *Condition* (Real vs Sham for experimental group; Pseudo-Real vs Pseudo-Sham for control group) as a repeated-subjects factor. In all

analyses, there was no significant main effect of *Condition*, regardless of whether the dependent measure was average number of switches ($_{\text{experimental}}F(1,38) = .01$, $p = .91$, $\eta_p^2 < .001$; $_{\text{control}}F(1,23) = .05$, $p = .83$, $\eta_p^2 = .002$) or average size of clusters ($_{\text{experimental}}F(1,38) = .28$, $p = .60$, $\eta_p^2 = .01$; $_{\text{control}}F(1, 23) = .03$, $p = .86$, $\eta_p^2 = .001$).

We combined results from all experiments and participant groups to carry out three mixed factor ANOVA, with *Condition* (Real vs Sham for experimental group; Pseudo-Real vs Pseudo-Sham for control group) and *Fluency Task* (Phonetic vs. Semantic) as within-participants factors, and *Group* (Control vs Experimental) as a between-participants factor. In each analyses, we included average number of switches, average number of clusters, or average cluster size as the dependent measure.

Average number of switches. There was a significant main effect of *Fluency Task* ($F(1,61) = 66.83$, $p < .001$, $\eta_p^2 = .52$), with a greater number of switches in semantic fluency compared to phonemic fluency (16.6±.4 vs 20.8±.5, respectively). There was a significant main effect of *Group* ($F(1,61) = 6.5$, $p = .01$, $\eta_p^2 = .10$), with a greater number of switches seen in the experimental compared to control group (19.8±.5 vs 17.7 ± .7, respectively). There was no interactions for *Condition* by *Group* ($F(1,61) = .05$, $p = .82$, $\eta_p^2 = .001$), *Condition* by *Task* ($F(1,61) = .60$, $p = .44$, $\eta_p^2 = .01$), but there was a significant interaction for *Task* by *Group* ($F(1,61) = 4.68$, $p = .03$, $\eta_p^2 = .07$). Crucially, there was no significant three-way interaction for *Condition* by *Group* by *Task* ($F(1,61) = .44$, $p = .51$, $\eta_p^2 = .01$).

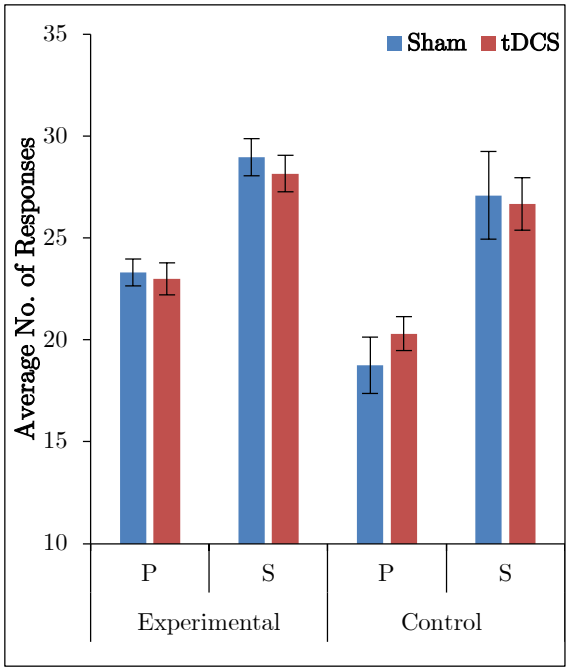
Average cluster size. There was a significant main effect of *Fluency Task* ($F(1,61) = 61.11$, $p < .001$, $\eta_p^2 = .50$), with an expected greater average cluster size in semantic fluency compared to phonemic fluency (2.3±.1 vs 3.5±.1, respectively). There was a significant main effect of *Group* ($F(1,61) = 15.3$, $p < .001$, $\eta_p^2 = .2$), with a greater average cluster size produced by the experimental compared to control group (3.2±.1 vs 2.6±.1, respectively). There was no significant interactions, including *Condition* by *Group* ($F(1,61) = .23$, $p = .80$, $\eta_p^2 = .001$), *Condition* by *Task* ($F(1,61) = .15$, $p = .70$, $\eta_p^2 = .002$), *Task* by *Group* ($F(1,61) = .16$, $p = .69$, $\eta_p^2 = .003$), and *Condition* by *Group* by *Task* ($F(1,61) = .15$, $p = .70$, $\eta_p^2 = .002$).

4.3.2.2 Subgroup analysis

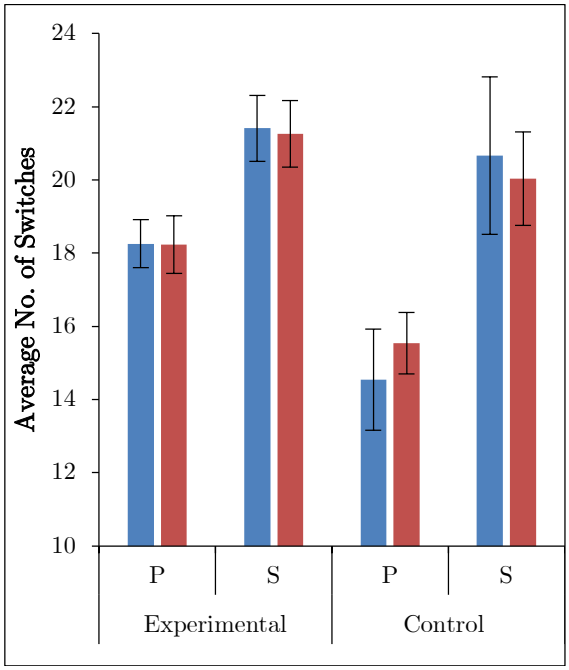
Effect of working memory span, reward sensitivity and switching ability. Figure 4.2 shows performance under sham and anodal tDCS with respect to variation in working memory span, motivation and switching ability. We carried out a series of mixed repeated ANOVAs, which included *Condition* (Real vs. Sham) as repeated-subjects factors, *Span, BAS-R* or *Switching* (High versus Low) as a between-subjects factor. We do not include control data here because the experimental group is of interest. The results showed a significant main effect of *Span* ($F(1,37) = 5.94$, $p = .02$, $\eta_p^2 = .14$), *BAS-R* ($F(1,37) = 5.17$, $p = .03$, $\eta_p^2 = .12$), and *Switching* ($F(1,37) = 26.50$, $p < .001$, $\eta_p^2 = .42$), and no significant interactions for *Condition* by *Span* ($F(1,37) = .10$, $p = .76$, $\eta_p^2 = .003$), by *BAS-R* ($F(1,37) = 2.00$, $p = .17$, $\eta_p^2 = .05$), or by *Switching* ($F(1,37) = 3.45$, $p = .07$, $\eta_p^2 = .01$).

VERBAL FLUENCY – GROUP ANALYSIS

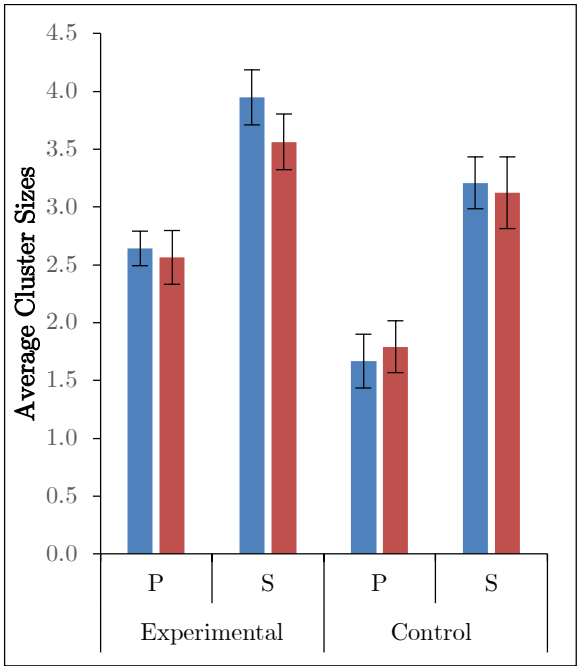
a. Overall Performance



b. Average No. of Switches



c. Average Size of Clusters



Legend: P = Phonemic Fluency; S = Semantic Fluency

Fig. 4.1 – Performance across sham and tDCS conditions and participant groups in terms overall average no. of responses (A); average no. of switches (B), and average size of clusters (C). Error Bars reflect Standard Error.

VERBAL FLUENCY – SUBGROUP ANALYSIS

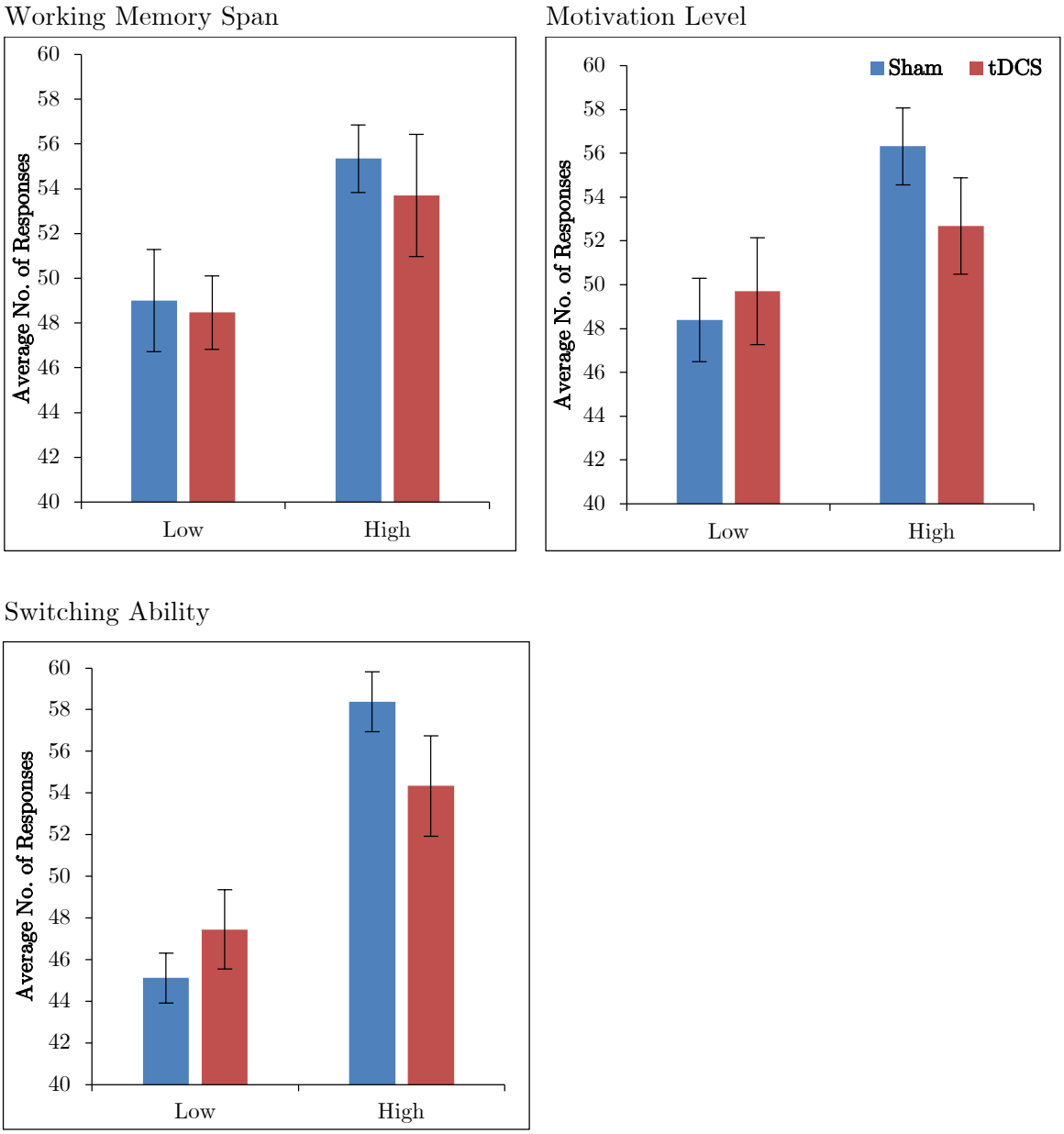


Fig. 4.2 – Performance across stimulation conditions, with participant subdivided by Working Memory Span; Motivation Level; and Switching Ability. Error Bars reflect Standard Error.

4.3.3 Results for Experiment 2: Probe Tasks

4.3.3.1 Group analysis

Figure 4.3 show the average performance across participant groups for each probe task, probe type and for each stimulation condition. It is clear that our tasks were sensitive to our manipulation, which we assume is related to executive selection. For instance, relative to negative (neutral) probe trials, longer reaction times and higher percentage errors were observed on lures (e.g., non-recent- and recent-probes in recent-probe; negative-combined plus associated, negative-associated and negative-combined in semantic-associated probe; for similar results, Jonides & Nee, 2006). A similar pattern was seen when comparing positive with negative neutral probe trials. This last finding was not in line with the literature, but not completely unexpected (see Jonides & Nee, 2006). Participants would naturally apply the more effortful executive selection strategy to all probes they had seen before or were similar to the target list because these could potentially be a lure. Importantly, tDCS did not systematically modify performance.

We compared the difference between stimulation conditions separately for data from each participant group and probe task. We carried out a series of individual one-way ANOVAs, with *Condition* (Sham vs tDCS for experimental group; Pseudo-Sham vs Pseudo-tDCS for control group) as a within-participants factor, and reaction times or percentage errors as the dependent measure. The results showed no main effect of *Condition* in recent-probe for the experimental group ($_{RTs}F(1,19) < .001$, $p = .99$, $\eta_p^2 < .001$; $_{ACC}F(1,19) = .004$, $p = .95$, $\eta_p^2 < .001$) and control group ($_{RTs}F(1,23) = .10$, $p = .75$, $\eta_p^2 = .004$; $_{errors}F(1,23) < .001$, $p = .99$, $\eta_p^2 < .001$) or the semantic-associated probe for the experimental group ($_{RTs}F(1,18) = .004$, $p = .95$, $\eta_p^2 < .001$; $_{errors}F(1,18) = .62$, $p = .44$, $\eta_p^2 = .03$) and control group ($_{RTs}F(1,23) = 1.78$, $p = .20$, $\eta_p^2 < .07$; $_{errors}F(1,23) = .28$, $p = .61$, $\eta_p^2 = .01$).

We carried out mixed factor ANOVAs combing data from both participant groups, with *Condition* (Sham vs tDCS for experimental group; Pseudo-Sham vs Pseudo-tDCS for control group) and *Probe Conditions* (Negative vs Positive) as within-participants factors, and *Group* (Control vs Experimental) as a between-participants factor.

For the *recent-probe*, we found a significant main effect of *Probe Condition* ($_{\text{RTs}}F(3,126) = 11.62$, $p < .001$, $\eta_p^2 = .23$; $_{\text{errors}}F(3,126) = 44.70$, $p < .001$, $\eta_p^2 = .52$), and *Group* for reaction times but not errors ($_{\text{RTs}}F(1,42) = 4.19$, $p = .05$, $\eta_p^2 = .09$; $_{\text{errors}}F(1,42) = .003$, $p = .96$, $\eta_p^2 < .001$), with experimental being slower the control group (1032±53 vs 886±49). The *Probe Condition* by *Group* interaction was significant for errors but not reaction times ($_{\text{RTs}}F(3,126) = .72$, $p = .54$, $\eta_p^2 = .02$; $_{\text{errors}}(F(3,126) = 9.91$, $p < .001$, $\eta_p^2 = .19$). Crucially, there was no significant interaction for *Condition* by *Group* ($_{\text{RTs}}F(1,42) = .03$, $p = .77$, $\eta_p^2 = .001$; $_{\text{errors}}F(1,42) = .002$, $p = .96$, $\eta_p^2 < .001$), *Condition* by *Probe Condition* ($_{\text{RTs}}F(3,126) = .49$, $p = .69$, $\eta_p^2 = .01$; $_{\text{errors}}F(3,126) = 1.24$, $p = .30$, $\eta_p^2 = .03$), or *Condition* by *Group* by *Probe Condition* ($_{\text{RTs}}F(3,126) = .53$, $p = .66$, $\eta_p^2 = .01$; $_{\text{errors}}F(3,126) = .38$, $p = .77$, $\eta_p^2 = .01$).

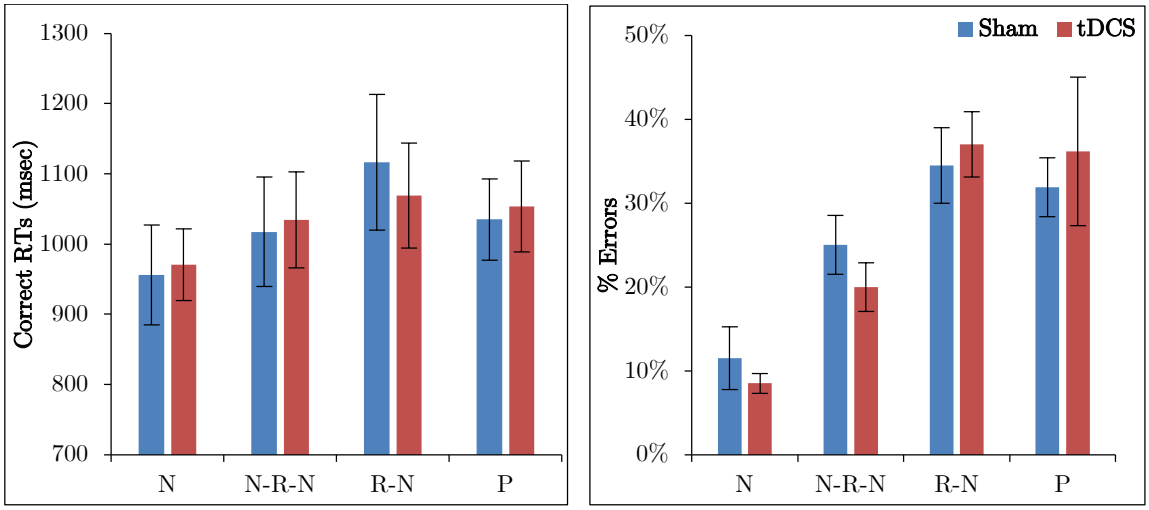
For the *semantic-associated probe*, we found a significant main effect of *Probe Condition* ($_{\text{RTs}}F(5,205) = 5.93$, $p < .001$, $\eta_p^2 = .13$; $_{\text{errors}}F(5,205) = 24.63$, $p < .001$, $\eta_p^2 = .38$), and*Group* ($_{\text{RTs}}F(1,41) = 4.41$, $p = .04$, $\eta_p^2 = .10$; $_{\text{errors}}F(1,41) = 39.18$, $p < .001$, $\eta_p^2 = .49$), with control participants being significantly faster and more error prone than experimental participants (869±25 vs 798±23 msec; 11±2 vs 24±1%, respectively). There was no significant interaction for *Probe Condition* by *Group* for percentage errors but not reaction times ($_{\text{RTs}}F(5,205) = 1.28$, $p = .28$, $\eta_p^2 = .03$; $_{\text{errors}}F(5,205) = 3.13$, $p = .01$, $\eta_p^2 = .07$), but there was not significant interaction for *Condition* x *Group* ($_{\text{RTs}}F(1,41) < .001$, $p = .99$, $\eta_p^2 < .001$; $_{\text{errors}}F(1,41) = .01$, $p = .94$, $\eta_p^2 = .00$), *Condition* by *Probe Condition* ($_{\text{RTs}}F(5,205) = .39$, $p = .85$, $\eta_p^2 = .01$; $_{\text{errors}}F(5,205) = .71$, $p = .62$, $\eta_p^2 = .02$), or *Condition* by *Group* by *Probe Condition* ($_{\text{RTs}}F(5,205) = 1.34$, $p = .34$, $\eta_p^2 = .03$; $_{\text{errors}}F(5,205) = 2.12$, $p = .06$, $\eta_p^2 = .05$).

PROBES TASKS – GROUP ANALYSIS

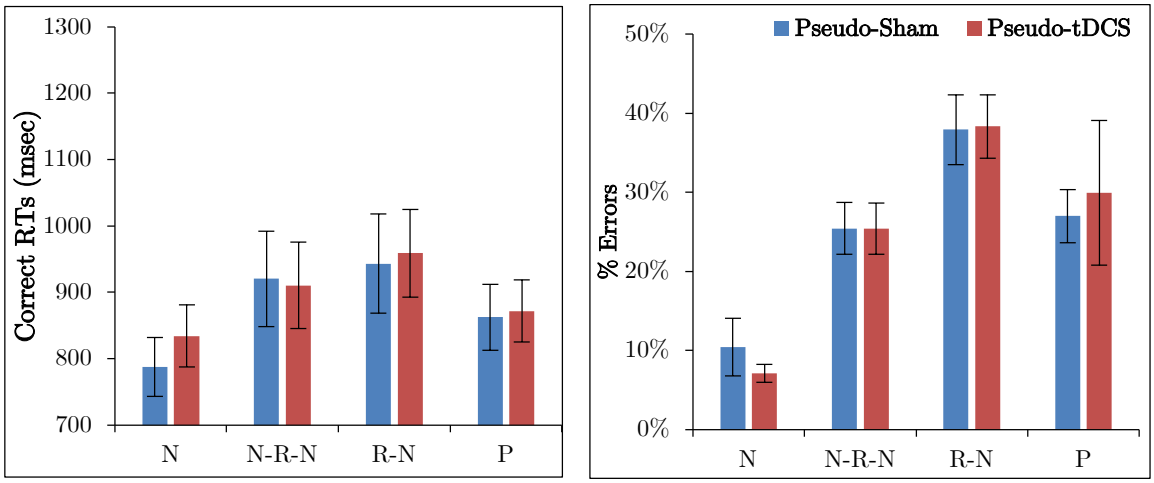
Reaction Times (msec) **Percentage Errors**

a. Recent-Probe

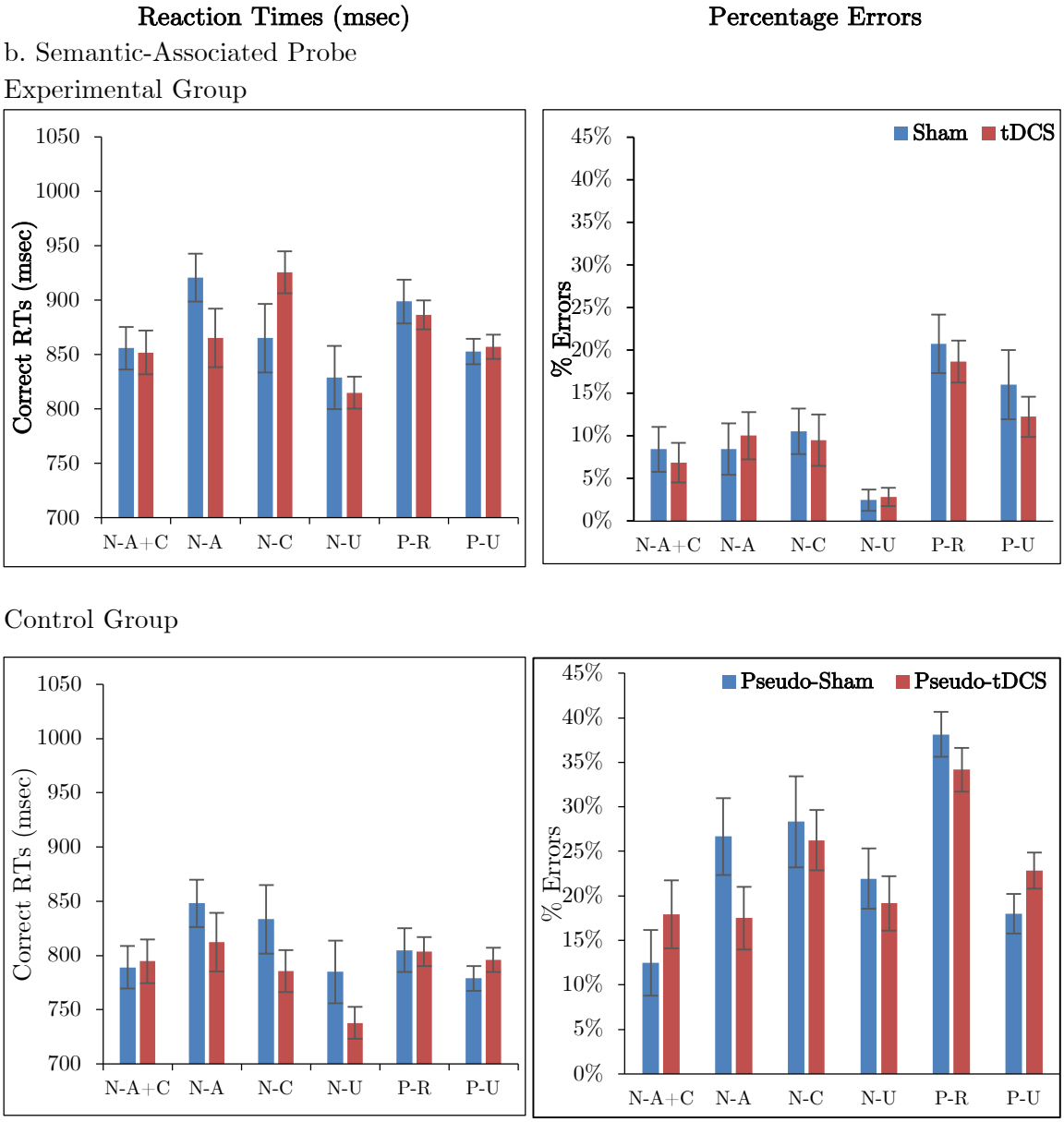
Experimental Group



Control Group



Legend: N = Negative; N-R-N = Non-Recent-Negative; R-N = Recent-Negative, P = Positive



Legend: N-A+C = Negative-Associated plus Combined; N-A = Negative-Associated; N-C = Negative-Combined; Negative-Unrelated; P-R = Positive-Related; P-U = Positive-Unrelated

Fig. 4.3 – Average correct RTs (msec) and percentages errors for each probe task (recent-probe, panel a; semantic-associated probe, panel b) across participant groups, probe conditions, and stimulation conditions. Error Bars reflect Standard Error.

Effect of stimulation on aggregated interference. We thought the effects of tDCS might be detectable if the interference effect from lures were aggregated across lure conditions. For recent-probe, we calculated aggregated interference as the difference between negative probes and the average of recent-negative and non-recent-probes. For the semantic-associated probe, we calculated aggregated interference as the difference between negative unrelated probes with the average of negative-associated and -combined. The aggregated interference effects across probe tasks, participant groups and stimulation conditions are presented in Figure 4.4 It is clear that tDCS had no systematic effect on the magnitude of aggregated interference.

We carried out separate mixed factor ANOVAs, with *Condition* as a within-participants factor, *Group* (Experimental vs Control) as a between-participants factor, and *aggregated interference* as a dependent measure. For *recent-probe*, there was no main effect of *Condition* ($_{RTs}F(1,42) = 1.34$, $p = .25$, $\eta_p^2 = .03$; $_{errors}F(1,42) = 1.35$, $p = .25$, $\eta_p^2 = .03$) and *Group* ($_{RTs}F(1,42) = .35$, $p = .56$, $\eta_p^2 = .01$; $_{errors}F(1,42) = 2.35$, $p = .13$, $\eta_p^2 = .05$), and no significant *Condition* by *Group* interaction ($_{RTs}F(1,42) = .05$, $p = .83$, $\eta_p^2 = .001$; $_{errors}F(1,42) = .01$, $p = .92$, $\eta_p^2 < .001$). For *semantic-associated probe*, there was no main effect of *Condition* ($_{RTs}F(1,41) = .42$, $p = .52$, $\eta_p^2 = .01$; $_{errors}F(1,41) = .02$, $p = .90$, $\eta_p^2 = .00$) or a main effect of *Group* ($_{RTs}F(1,41) = .14$, $p = .71$, $\eta_p^2 = .00$; $_{errors}F(1,41) = 2.87$, $p = .10$, $\eta_p^2 = .07$), and no significant *Condition* x *Group* interaction ($_{RTs}F(1,41) = .02$, $p = .90$, $\eta_p^2 = .00$; $_{errors}F(1,41) = .02$, $p = .90$, $\eta_p^2 = .00$)

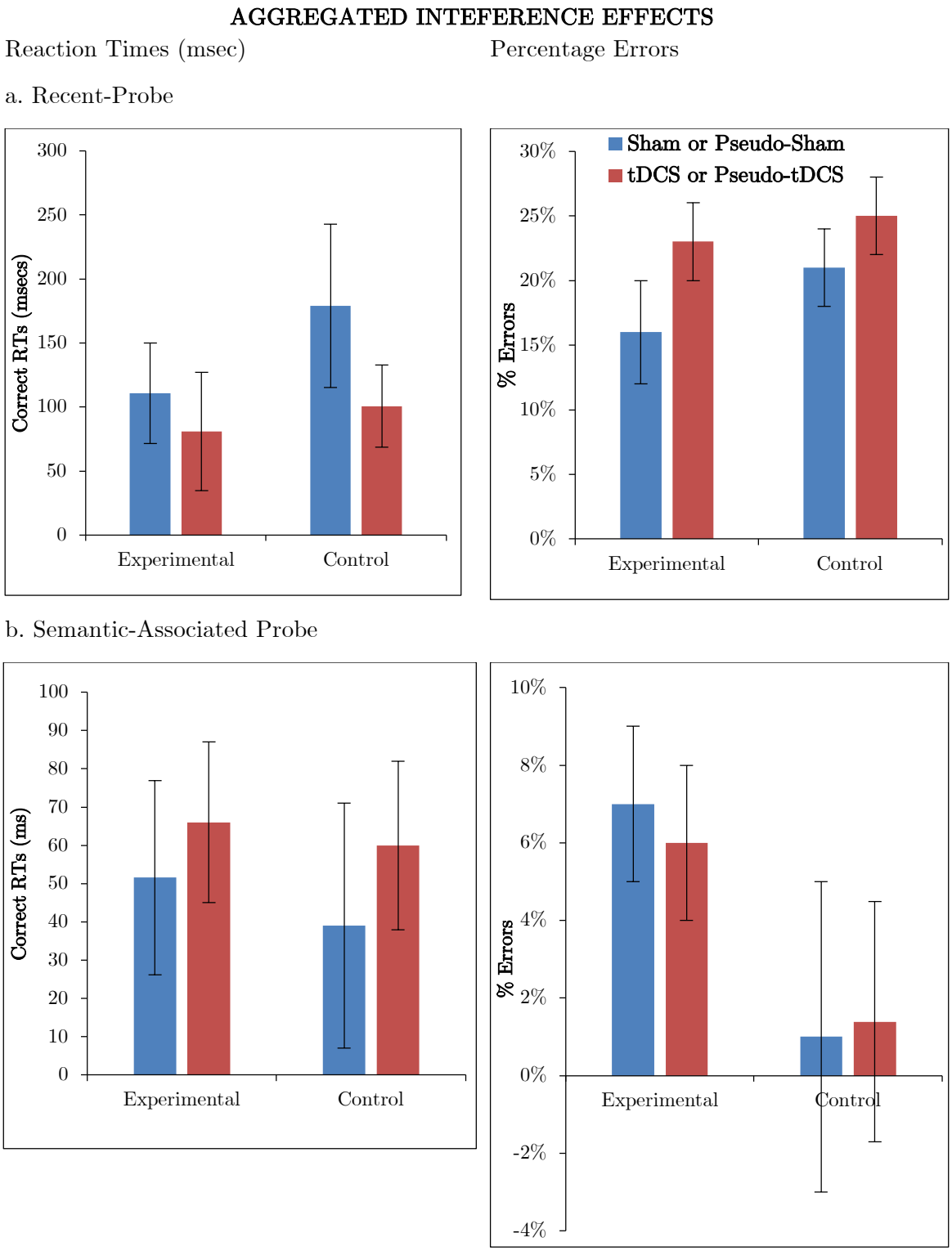


Fig. 4.4 – Aggregated interference across probe tasks (recent-probe, panel a; semantic-associated probe, panel b) stimulation conditions and participant groups. Interference measured as difference between negative probes and the average of recent-negative and non-recent-probes or as difference between negative unrelated probes with the average of negative-associated and –combined. Error Bars reflect Standard Error.

4.3.3.2 Subgroup analysis

Effect of working memory span, reward sensitivity and interference control. Figure 4.5 shows performance under sham and anodal tDCS with respect to variation in working memory span, motivation and switching ability. We carried out a series of mixed factors ANOVAs, which included *Condition* (Real vs. Sham) as repeated-subjects factors, *Span, BAS-R* or *Interference Control* (High versus Low) as a between-subjects factor, and *aggregated interference* as the dependent measure. We carried out a series of mixed repeated ANOVAs, which included *Condition* (Real vs. Sham), and *Probe Conditions* (Negative vs Positive) as within-participants factors, *Span, BAS-R* or *Interference Control* (High versus Low) as a between-subjects factor. We do not include control data here because the experimental group is of interest.

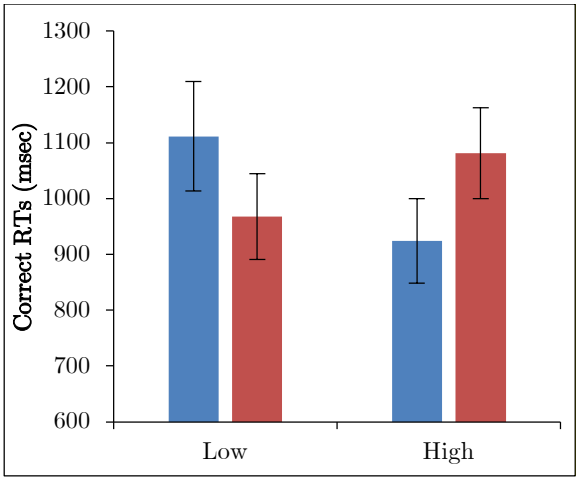
For *recent-probe*, the results showed significant main effects of *Span* ($_{RTs}F(1,18) = 9.22$, $p = .01$, $\eta_p^2 = .20$; $_{errors}F(1,18) = .01$, $p = .91$, $\eta_p^2 = .001$), but not *BAS-R* ($_{RTs}F(1,18) = .06$, $p = .81$, $\eta_p^2 = .003$; $_{errors}F(1,18) = .16$, $p = .69$, $\eta_p^2 = .01$), and *Interference Control* ($_{RTs}F(1,18) = .36$, $p = .56$, $\eta_p^2 = .02$; $_{errors}F(1,18) = .03$, $p = .86$, $\eta_p^2 = .002$). There was no significant interactions for *Condition* by *Span* ($_{RTs}F(1,18) = .01$, $p = .99$, $\eta_p^2 < .001$; $_{errors}F(1,18) = .58$, $p = .46$, $\eta_p^2 = .03$), by *BAS-R* ($_{RTs}F(1,18) = .01$, $p = .94$, $\eta_p^2 < .001$; $_{errors}F(1,18) = .79$, $p = .39$, $\eta_p^2 = .04$), and by *Interference Control* interaction ($_{RTs}F(1,18) = .04$, $p = .84$, $\eta_p^2 = .002$; $_{errors}F(1,18) = .17$, $p = .69$, $\eta_p^2 = .01$). For semantic-associated probe, the results showed a main effect of *Span* for reaction times only ($_{RTs}F(1,17) = 9.13$, $p = .01$, $\eta_p^2 = .35$; $_{errors}F(1,17) = .77$, $p = .39$, $\eta_p^2 = .04$), but not by *BAS-R* ($_{RTs}F(1,17) = .32$, $p = .58$, $\eta_p^2 = .02$; $_{errors}F(1,17) = .25$, $p = .62$, $\eta_p^2 = .02$), an a significant main effect by *Interference Control* for errors only ($_{RTs}F(1,17) = .32$, $p = .58$, $\eta_p^2 = .02$; $_{errors}F(1,17) = 7.84$, $p = .01$, $\eta_p^2 = .32$). There was no significant interactions for *Condition* by *Span* interaction ($_{RTs}F(1,17) = .67$, $p = .42$, $\eta_p^2 = .04$; $_{errors}F(1,17) = .51$, $p = .49$, $\eta_p^2 = .03$), *BAS-R* ($_{RTs}F(1,17) = 1.80$, $p = .20$, $\eta_p^2 = .10$; $_{errors}F(1,18) = .40$, $p = .54$, $\eta_p^2 = .02$), nor *Interference Control* interaction ($_{RTs}F(1,17) = 1.80$, $p = .20$, $\eta_p^2 = .10$; $_{errors}F(1,17) = .31$, $p = .59$, $\eta_p^2 = .02$).

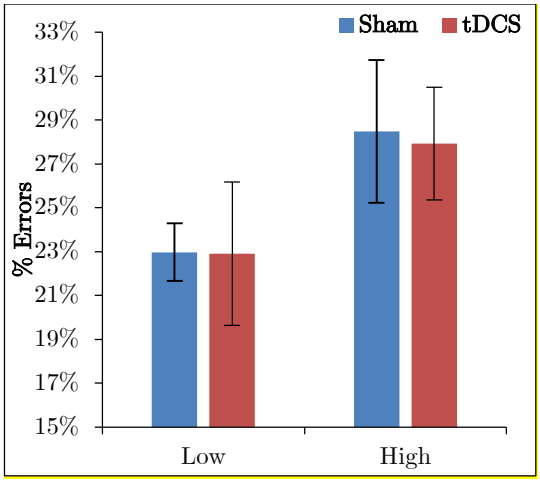
PROBE TASK – SUBGROUP ANALYSIS

Reaction Times (msecs) | Percentage Errors

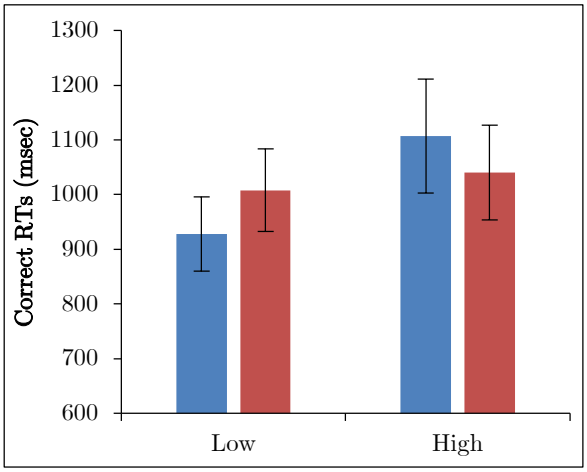
a. Recent-Probe

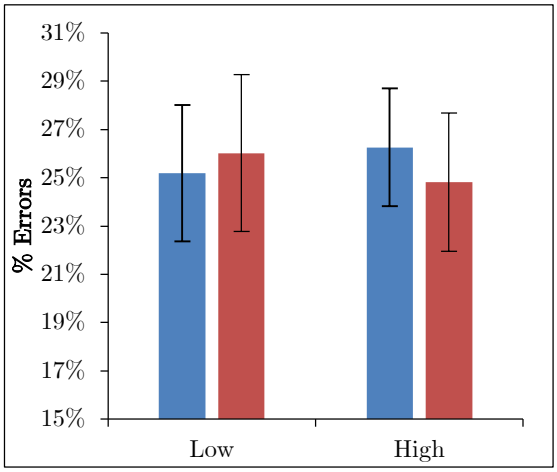
Working Memory Span



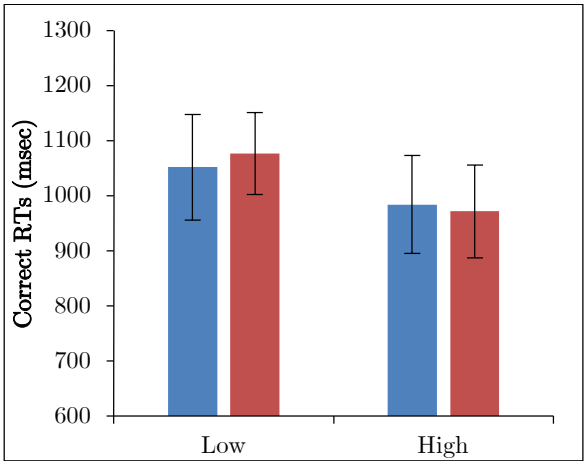


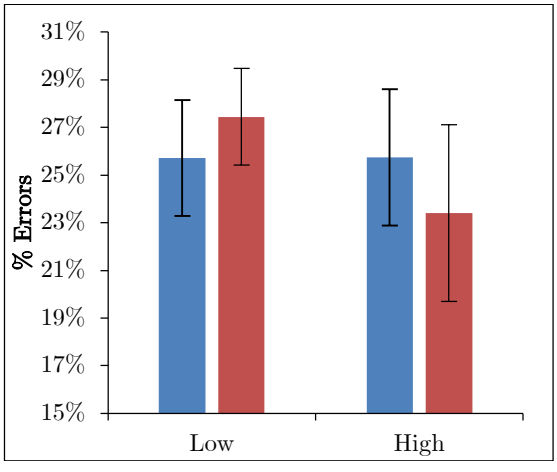
Motivation Level





Interference Control





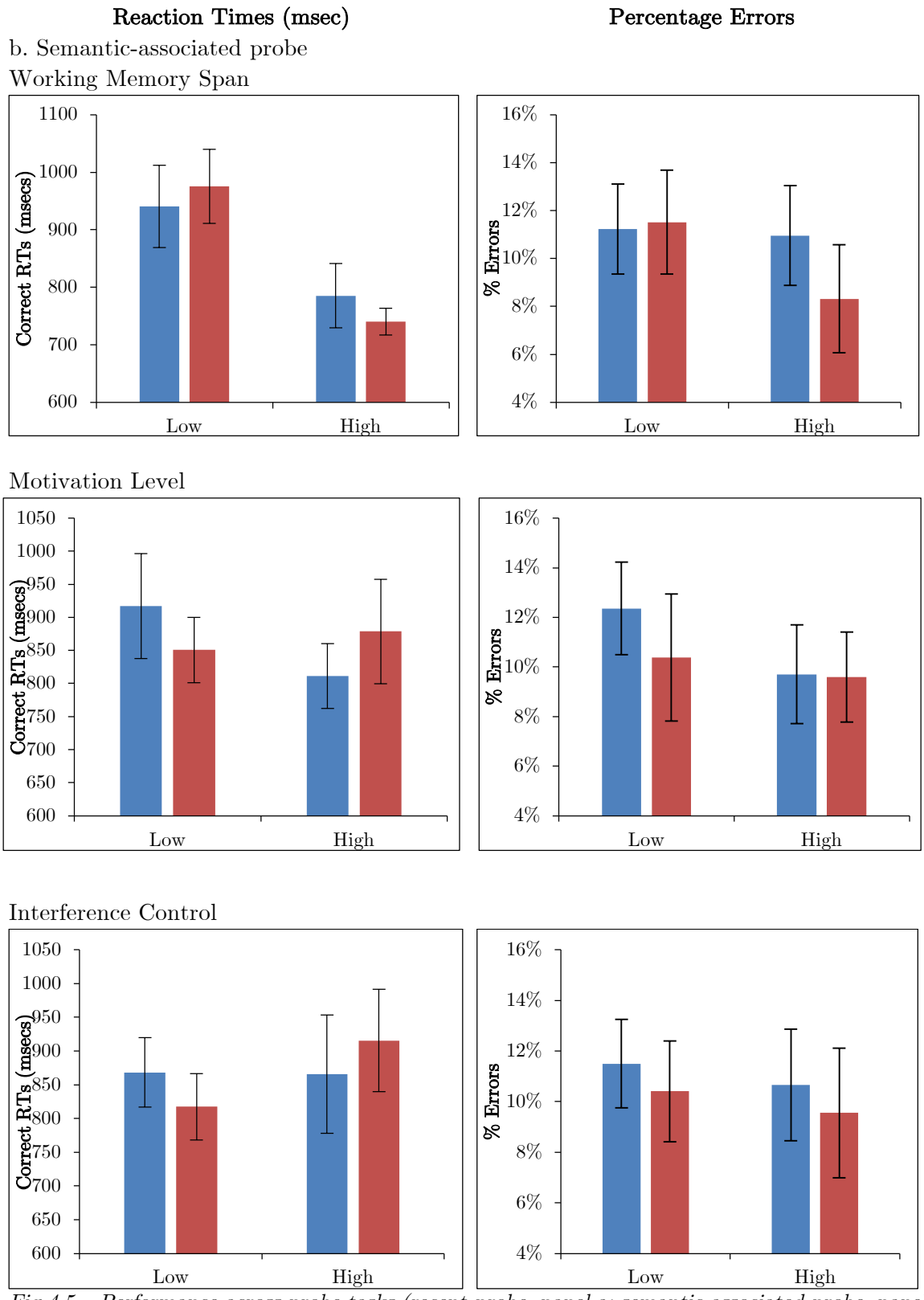


Fig 4.5 – Performance across probe tasks (recent-probe, panel a; semantic-associated probe, panel b) and stimulation conditions, with participants divided by Working Memory Span; Motivation Level; and Switching Ability. Error Bars reflect Standard Error.

4.4 Discussion

When modulating cognition in healthy participants, single sessions of anodal tDCS produces unreliable results (Horvath et al., 2015a; Westwood & Romani, 2017). We carried out a fresh series of experiments to see whether tDCS could modify performance on tasks that probe executive selection abilities, namely verbal fluency and probe tasks, using conventional stimulation parameters – i.e., 1.5mA anodal tDCS applied for 25 mins during task performance. In line with previous evidence, we expected that performance would be enhanced, given the role the LIFG plays in executive selection (Badre, 2008; Badre & Wagner, 2007; Shao et al., 2014). We also accounted for the fact that the tDCS effect can be confounded by baseline and task mediated variation in executive selection ability. Thus, we considered overall performance at the group level, looking at switching in fluency and responses to lures in probe tasks, and additional analysis which examined absolute effects of tDCS and performance on difficult items (see Appendix 4.5). We then sought to see how performance fared when participants were divided into subgroups by working memory span, motivation levels, and switching ability or control of interference. Despite our efforts, we found no systematic effect of stimulation.

Our results are in line with the inconsistent outcomes across studies attempting to modulate verbal fluency and working memory tasks in healthy participants with single sessions of anodal tDCS, as adumbrated in the introduction. Such differences in outcomes could reflect the variation in stimulation protocols across studies, but there does not seem to be a consistent association between a particular protocol and positive results. Appendix 4.1 shows a comparison between the protocols used in the studies reported in Experiments 1 and 2 and similar published studies. The protocol adopted by Marshall et al. (2005) is perhaps the most salient exception to convention, targeting the left/right dlPFC simultaneously with intermittent tDCS. Our study differs in the site and timing of stimulation, but our decision to target the LIFG using online stimulation should have increased, not decreased, the effectiveness of tDCS. Neuroimaging studies report greater cortical excitation during rather than after tDCS administration (Martin, Liu, Alonzo, Green & Loo, 2014; Stagg et al. 2013; Rae et al., 2013), and online effects are thought to operate on neuronal populations activated by the task (Bikson & Rahman, 2013; Lapenta et al., 2013; Pisoni et al., 2017). In other key

aspects (e.g., current density, site of reference electrode, stimulation duration), we fall well within the range used by previous studies.

Differently from us, other studies targeted the left temporal lobe. Three studies have reported positive effects with left temporal anodal tDCS, with a significant reduction in false memories in the DRM task (Boggio et al., 2009; Díez et al., 2017) and better performance in a recent-probe without lures (Pisoni et al., 2015). The left temporal lobe is important in lexical access, which is required for our tasks. However, for fluency and probe tasks top-down frontal selection mechanism are also likely to be engaged since in these tasks there is a special need to move from one lexical/semantic field to another (fluency tasks) and/or to inhibit distractors (probe tasks). In fact, neuroimaging studies show that frontal regions work in concert with temporal regions to mediate performance in these tasks (e.g., Badre, 2008; Biesbroek et al., 2016). One may assume, therefore, that temporal stimulation would improve performance on verbal fluency and item recognition tasks, by facilitating lexical retrieval and/or maintenance of task relevant information. Boosting selection control mechanisms, however, should also have a positive effect.

Clearly, the null effects we report may still be a result of the failure to use *optimal combination* of parameters, but the fact is that conditions in which reliable effects of tDCS can be measured have not been established, at least within conditions covered by our study (e.g., fluency, working memory, healthy participants, and one session of anodal tDCS). A good way to test the possibility that null effects were due to protocol differences—and to elucidate conditions in which tDCS can operate optimally—is to conduct direct replications of studies that report positive effects. In yet unpublished work, we failed to replicate Cattaneo et al. (2011), which reported a large positive effect on semantic and phonemic fluency tasks after anodal tDCS was applied to the LIFG. We encourage others to confirm the reliability of previous findings by way of direct or conceptual replication.

Lack of power could be another reason for null outcomes because it reduces the likelihood of finding a true effect, if one exists. The sample sizes used in our studies (n = 19 and 20) are relatively small, but consistent with previous studies that found positive effects on fluency and working memory (see Price et al., 2015; Mancuso et al., 2016). Our aggregated

sample size from Experiment 1 (n = 39), gave us good power to detect a large (0.8) and medium (0.5) effect size (1-β = 0.99 and 0.86, respectively), but we had limited power to detect a small effect size (0.2; 1-β = 0.22, α = 0.05). A meta-analysis is an ideal means to evaluate effects across individual studies that are underpowered. Mancuso et al. (2016) reported results indicating that effects on working memory tasks are generally small or non-significant even with a large sample of 471 (Hedges' g = ˜0.2, see left dlPFC analysis). Price et al. (2015), however, reported a significant mean effect size of roughly 0.5 (Hedges' g) with a large sample (n = 119) across studies measuring verbal fluency and word learning tasks. In yet unpublished work, we pooled data from Price et al., and several studies published since, including our data from Experiment 1. The results showed that with a sample of roughly 230 participants, anodal tDCS significantly improved fluency performance. Still, this effect was more moderate than reported by Price et al. (2015; roughly 0.3, Hedges' g), and potentially inflated by exceptionally large treatment effects from underpowered studies. Thus, it remains to be seen if, for fluency tasks, tDCS effects are stable for properly powered studies.

Finally, we used the BIS/BAS motivational scale to assess whether the effect of tDCS may interact with reward sensitivity. One possible alternative approach would be to use a measure that is directly related to the task. The goal of the paper, however, was to identify general moderators that may serve to refine conventional protocols, and trait reward sensitivity is an ideal candidate because high BAS scores are associated with performance on working memory and cognitive control functions (for reviews, see Gray and Burgess, 2004; Jonides and Nee, 2006; Savine et al., 2010; Fino et al., 2014). More importantly, however, we chose the BAS scale because it has previously been used to identify responders to tDCS modulation (see Metuki et al., 2012; Sela et al., 2012). Another possibility is that we manipulate the extent to which a task is rewarding. Only one study—to our knowledge—has investigated this, and found that a financial incentive improved the effect of anodal tDCS (p = 0.04, see Jones et al., 2015). It might be that an external motivation to do well on a task can boost the facilitatory effect of tDCS.

Conclusions

We focused on single sessions of tDCS in healthy individuals and found negative results. We do not want, however, to dismiss possible stronger effects of tDCS in other conditions. Cortical excitability in healthy brains is potentially already at optimal levels, meaning that null effects may be due to ceiling effects. More reliable effects may be seen when anodal stimulation is compared with cathodal stimulation which should decrease performance (for review, see Jacobson et al., 2012; Horvath et al., 2015). It remains likely, however, that, in healthy brains, homeostatic mechanisms may reduce or even nullify the effect of tDCS in order to maintain stable network activity (Krause and Kadosh, 2014). Positive effects may be more likely in participants with pathological or reduced levels of excitability. For example, more consistent effects of tDCS have been reported in patients with aphasia (see Monti et al., 2013; de Aguiar et al., 2015; Elsner et al., 2015; Sandars et al., 2015; Shah-Basak et al., 2015; Cappon et al., 2016; Crinion, 2016). Stronger effects may also occur in tasks where processes and representations are not yet stable, such as in the case of learning. Lastly, positive effects may be more likely when tDCS is applied across repeated sessions, thereby allowing for effects to accumulate (Alonzo et al., 2012; Meinzer et al., 2014). Indeed, a number of studies have shown enhanced learning following repeated stimulation even in normal participants (Flöel et al., 2008; Dockery et al., 2009; Reis et al., 2009; Kadosh et al., 2010; Meinzer et al., 2014). Our study should encourage further studies to establish the conditions where tDCS effects are stronger and/or more reliable.

Chapter 5: Replication and meta-analysis investigating the effects of tDCS on verbal fluency

5.1 Overview

The null effects reported in Chapter 4 prompted further investigation to see whether tDCS can successfully modulate verbal fluency performance. I therefore carried out a systematic-review and meta-analysis and a direct replication to respectively quantify and re-evaluate previous reports of positive effects.

5.2 Introduction

Findings from studies measuring the effects of single session anodal tDCS show inconsistent outcomes on verbal fluency tasks in healthy participants (see Appendix 5.1, where I provide a summary of main findings). Attempts to quantify the efficacy of tDCS on fluency tasks via quantitative reviews and meta-analyses unfortunately result in equally inconsistent outcomes. It is commonly cited that this variation may have arisen from differences in stimulation protocols, rather than the poor efficacy of tDCS. However, reviews of effects in word production and working memory tasks show that lack of power and the presence of publication bias are potential culprits (see Mancuso et al., 2016; Westwood & Romani, 2017). I therefore decided to re-evaluate the effectiveness of tDCS on verbal fluency tasks in a meta-analysis and a direct replication of a previous study to contribute to this question.

Previous attempts to collate findings and quantify the effect of tDCS on verbal fluency tasks produce conflicting results. In a large scale quantitative review, Horvath et al. (2015a) quantified the effects of single sessions of tDCS applied to healthy participants across 80 measures of cognition. None of the 59 analyses carried out produced a significant result, of which 3 different analyses pooled effects from 5 verbal fluency studies. However, this review has been criticised for inclusion criteria that may have reduced the chances of a significant outcome. For example, to reduce the impact of idiosyncratic methodologies (e.g., tDCS devices, stimulation protocol), the authors excluded task/outcome measures that had not been used by at least two separate research groups, thereby reducing the number of eligible

studies. It also meant that analyses typically pooled as few as two or three studies (see Price & Hamilton, 2015). Price et al. (2015) re-examined this data, focusing on studies using anodal tDCS on word learning and verbal fluency tasks. The authors decided to pool all studies together to boost power, as well as carrying out additional analyses to see if the effect of tDCS was moderated by the timing of stimulation. The result indicated that performance improved with anodal tDCS, either when pooling all studies together (n = 8), studies that applied offline tDCS (n = 4), or verbal fluency studies using offline tDCS (n = 3). Still, effects were small to moderate (roughly ~0.5 Hedges' g). However, more importantly, because effects were drawn from a small sample of underpowered studies, the summary effect was potentially inflated.

An underpowered study naturally increases the chances of finding a null effect. However, for a significant result the effect is likely to be an over estimation of the true effect size (for review, see Button et al., 2013), and publication bias (i.e., favouring the publication of positive over null effects) can keep a field with underpowered studies hostage to these large treatment effects. In Price et al. (2015), one can see that out of the eight included studies, only three reported positive effects, which were exceptionally large (one from a word learning study, ~0.8, Floel et al., 2008; two from verbal fluency studies, ~1.1; Cattaneo et al., 2011; ~.7, Meinzer et al., 2012), whilst the effect of offline tDCS on fluency arose from two small null effects (i.e., ~.2, Cerrutti & Schlaug, 2009; ~.3, Penolazzi et al., 2013a) and one large positive effect (~1.1, Cattaneo et al., 2011). Price et al. (2015) did not consider the impact of publication bias and/or study heterogeneity on the final outcome of their meta-analysis, despite the fact that sample sizes ranged from 10 to 20 participants. Previous meta-analysis have shown publication bias across studies using word production and working memory tasks (see Chapter 3; Mancuso et al., 2016), and, what is worse, there are very few attempts to replicate the large effects reported in previous studies (see Vannorsdal et al., 2016).

Thus, I carried out a systematic review and meta-analysis along with a direct replication of Cattaneo et al. (2011) in an attempt to quantify and re-evaluate the effect of tDCS in verbal fluency tasks. This investigation has several strengths over previous studies:

1. Since Price et al. (2015) several papers have been published investigating the effects of tDCS on verbal fluency tasks, which will boost the power of my analysis, and give a more precise effect size estimate.
2. The studies I include will likely be underpowered, so to account for inflated/chance effects in the summary effect size estimate, I will run a heterogeneity and publication bias analysis.
3. The effects of tDCS appears to be moderated by the type of fluency task, with larger effects seen on semantic versus phonemic. I will therefore carry out separate analyses to see which task is more amenable to the tDCS effect.
4. Finally, previous attempts to replicate Cattaneo et al. (2011) have failed, but these had significant shortcomings. I therefore carried out a stricter replication.

5.3. Experiment 1: A strict replication of Cattaneo et al. (2011)

Here, I conducted a strict direct replication of Cattaneo et al. (2011; hereafter referred to as Cattaneo et al.). In this study, ten participants were asked to perform a phonemic and semantic fluency task immediately after 2mA of anodal tDCS was applied to the LIFG. The results showed that performance across both fluency tasks was significantly enhanced. I am aware of one partially failed attempt to *reproduce* (Penolazzi et al., 2013a) and one completely failed attempt to *replicate* (Vannorsdal et al., 2016) findings reported by Cattaneo et al. Both of these studies are not strict replications of Cattaneo et al, and therefore fail to verify whether the protocol used by Cattaneo et al. is or is not effective.

Penolazzi et al. (2013a) aimed to compare the effectiveness of different electrode montages to modify semantic fluency performance. Participants were asked to perform semantic fluency three times: before stimulation (baseline measure), immediately after stimulation cessation, and 18 minutes after stimulation cessation. The authors administered 2mA of anodal tDCS for 20mins (same as Catteneo et al.), and targeted the LIFG using four different electrode montages. In one condition, Penolazzi et al. used the same montage as Cattaneo et al., placing the active electrode over the LIFG and the reference electrode over the contralateral supraorbital area (which they referred to as the *frontal montage*). The other three montages positioned the anode over the LIFG and partially over the left anterior

temporal region, with the reference electrode placed on either a) the right supraorbital area (*frontal-temporal montage*); b) the right homologue area (*bilateral montage*); or c) the right supraorbital area but with a 100cm^2 electrode (*unilateral montage*). The results showed that only the *frontal montage* enhance fluency performance relative to sham. However, this result differed to Cattaneo et al. For instance, fluency was enhanced only *18mins after* stimulation cessation (not *immediately after* stimulation cessation, as reported by Cattaneo et al.), and this enhancement disappeared when the authors analysed responses that were scored using a less stringent scoring criterion (this classed ambiguous responses as correct, such as "paraglider" or "skates" for the category "vehicles"). Since Penolazzi et al. (2013a) did not intend to perform a strict replication of Cattaneo et al., other differences in methodology and statistical analyses may have led to differences in outcomes:

1. Participants performed fluency before and after stimulation. We do not know what participants did during stimulation, which is relevant since activity during stimulation can confound the final outcome of tDCS (Holland & Crinion, 2012; Pisoni et al., 2017; Reis et al, 2009). In Cattaneo et al., participants watched a cartoon movie in an attempt to reduce variation in cortical activity during stimulation.
2. Why differences in scoring criteria would impact on the effect of tDCS was not explained or predicted by Penolazzi et al. (2013a) and may be an artefact of task instructions (which were not provided) or chance. Cattaneo et al. gave clear task instructions, excluded rule violations (e.g., proper names, repetitions of root words), and used categories that would generate unambiguous responses (e.g., Fruits, Animals, Car Brands, and Musical Instruments").
3. There is variation within participants (because participants vary at each time they were tested) and between participants (because different participants participated in each montage condition), which may have contributed to the effect of tDCS. For example, one would expect to see practice effects (e.g., more response at each time point), but for the sham group we see fewer responses at 18mins after stimulation compared to immediately after stimulation (see Fig. 2 reported by Penolazzi et al., 2013a). This could have easily masked the effect of tDCS immediately after stimulation cessation, or inflated this effect 18mins later. The within participant variation likely arose from task stimuli, namely the choice of semantic categories.

Penolazzi et al (2013a) counterbalanced pairs of semantic categories to appear equally at each time point participants were tested (i.e., before, immediately and 18mins after stimulation), and categories were selected because pilot data showed responses were equivalent across categories. Nonetheless, there is clear variation at each time point. Cattaneo et al. used fewer categories that produced equivalent responses (see Cattaneo et al., 2016) and a within participants design, both of which would have reduced variability in performance.

Vannorsdall et al (2016) is the only study conducted thus far that can be considered to be a direct replication of Cattaneo et al as far as the stimulation protocol is concerned. The results showed that tDCS failed to significantly influence performance on both phonemic and semantic fluency tasks. There are, however, problems with this study with regards to controlling practice effects, which could have masked the effect of tDCS. Fluency performance is particularly vulnerable to practice effects, especially if stimuli are repeated. Vannorsdall et al. (2016) used the same letters/categories in each stimulation session, and separated testing sessions by a relatively short period of time (as little as 24 hours), both of which increased the chances of practice effects. Cattaneo et al., instead, used different pairs of letters and categories in each stimulation session, which were counterbalanced across participants, and sessions were separated each testing session by 5 to 7 days. Cattaneo et al. (2016) elected to re-analyze data from Vannorsdall et al (2016), and found that significantly fewer words were produced in the first versus the second session (20.4 versus 22.6; $p = 0.001$).

5.3.1 Method

5.3.1.1 Participants

In Cattaneo et al., ten participants were recruited. In the present study, twenty-four undergraduates (18 female; 21 ± 2.59) were recruited from Aston University participated for course credits or financial reimbursement, and were assigned to the stimulation group in a semi-random fashion. All participants were right-handed and native English speakers, and all claimed to have normal or corrected to normal vision. We excluded volunteers with language impairments, history of migraine, headaches (frequent or severe), skin disorders (e.g., eczema), any adverse experience to previous tDCS, any history of epilepsy or stroke,

head/metal implants, any neurological disorders, and any volunteers who had participated in a tDCS or TMS study in the 6 months prior to the current study.

5.3.1.1 Overall Procedure

Participants received sham or real stimulation in two testing sessions, with each session separated by one week (plus or minus 1 day). In each session, participants were sat in front of a laptop computer in a quiet room, then given a screening form to sign (to conform eligibility for the study) followed by task instructions[11]. The instructions informed participants that they would receive tDCS stimulation for 20mins, during which time they would watch a short cartoon movie, then after stimulation they would perform a verbal fluency and semantic fluency task (specific task instructions were also given, which are provided below). Any questions from the participant were then answered, after which the researcher began applying the electrodes. The researcher then confirmed with the participant if they were still happy to continue before switching on the tDCS device and playing the cartoon movie (with no audio) on the laptop computer. Participants were reminded to watch the movie. The same cartoon movie was played for each participant and in each of the two testing sessions. The cartoon movie was stopped 2mins before the end of stimulation, and participants were told the fluency tasks would be performed in 2mins time. The first fluency task (e.g., phonemic) was started up to 1min after the end of stimulation, and the second fluency task (e.g., semantic) was started up to 1 min after the end of the first fluency task. Participants took between 3 to 5mins to complete both fluency tasks. The order of stimulation (sham or tDCS), and the pairing of stimulation with task stimuli for each fluency tasks in each session, was counterbalanced across participants. This procedure is a essentially an exact duplication of Cattaneo et al.

5.3.1.1 tDCS Procedure

In a single-blind procedure, sham/real stimulation was administered via a battery driven NeurConn DC-Stimulator using 35cm^2 saline-soaked electrodes, fixed to the scalp using

11In the first session, participants were given instructions about tDCS and consent form, before task instructions were given. Participants were asked if they had any questions about tDCS, and once these were answered, they were asked to sign the consent firm. Task instructions were then given.

elastic head straps. The anode electrode was placed over the LIFG, determined as the crossing point between the T3-Fz and F7-Cz, using the 10-20 EEG system. The cathode electrode was placed on the supraorbital area. For active stimulation, we applied a current intensity of 2mA for a duration of 20mins (with a 30s ramp up and down). Cattaneo et al. did not specify ramp duration, whilst Vannorsdall et al. used a 15s ramp up and down. We chose to use a ramp up and down duration of 30s since this was effective at retaining participant blinding (see Westwood et al., 2017; and above experiments). For the sham condition, the current was ramped up over 30s then immediately ramped down to zero over 30s. The sham procedure unavoidably differs – but only slightly – from Cattaneo et al., since they did not distinguish ramp duration from stimulation duration, instead they state, 'In both the real and the sham sessions, current was then turned off slowly over a few seconds (in the tDCS session after 20 min of stimulation, in the sham session after 30s of stimulation).' Nonetheless, the procedure we apply – i.e., ramping up and immediately ramping down the same current intensity used in active stimulation – is common, and achieves the intended goal, which is to induce the initial skin irritation (e.g., itching, tingling pins and needles) reported in active tDCS, thus making is difficult to distinguish with active stimulation, without modifying underlying cortical excitability (see Russo et al, 2013). The duration of stimulation (which is shown on the tDCS device display), was kept out of sight from the participant.

5.3.1.1 Verbal Fluency Task

We used the same instructions as in Experiment 1, 2a and 2b (see above; which were already modelled on Cattaneo et al.), and the same stimuli, which consisted of the letters *C, L, S*, *A* and the categories *Animals*, *Fruits*, *Super Market Items*, and *Musical Instruments*. Like Cattaneo et al., we presented different stimuli in each testing session to avoid practice effects. Unlike Cattaneo et al., however, we chose not to pair stimuli. We elected to counterbalance fully the order of letters/categories across testing sessions and each stimulation condition (sham versus real tDCS), for which we needed 24 participants (i.e., four stimuli conditions, hence: 1 x 2 x 3 x 4 = 24).

5.3.2 Results

Figure 5.4 presents the average number of correct responses across stimulation conditions and fluency tasks. There is a hint across both tasks that when administered anodal tDCS participants produced slightly more responses compared to sham (*M*±*SE:* phonemic, 26.5±1.4 versus 25.5±1.4; semantic, 38.1±1.4 versus 39.9±1.8). Like Cattaneo et al, we used a repeated factor ANOVA, with *Condition* (Real versus Sham) and *Task* (Phonemic versus Semantic) as a within-participants factors, and average correct average number of correct responses as a dependent measure. We found a significant main effect of *Task* ($F(1,11) = 50.01$, $p < .001$, $\eta_p^2 = .82$), but, more importantly, we found no main effect of *Condition* ($F(1,13) = 2.30$, $p = .16$, $\eta_p^2 = .17$), nor a significant *Condition* x *Task* interaction ($F(1,11) = .99$, $p = 0.35$, $\eta_p^2 = 0.08$). Thus, we failed to replicate the finding reported by Cattaneo et al., namely that anodal tDCS improved performance across both phonemic and semantic fluency tasks.

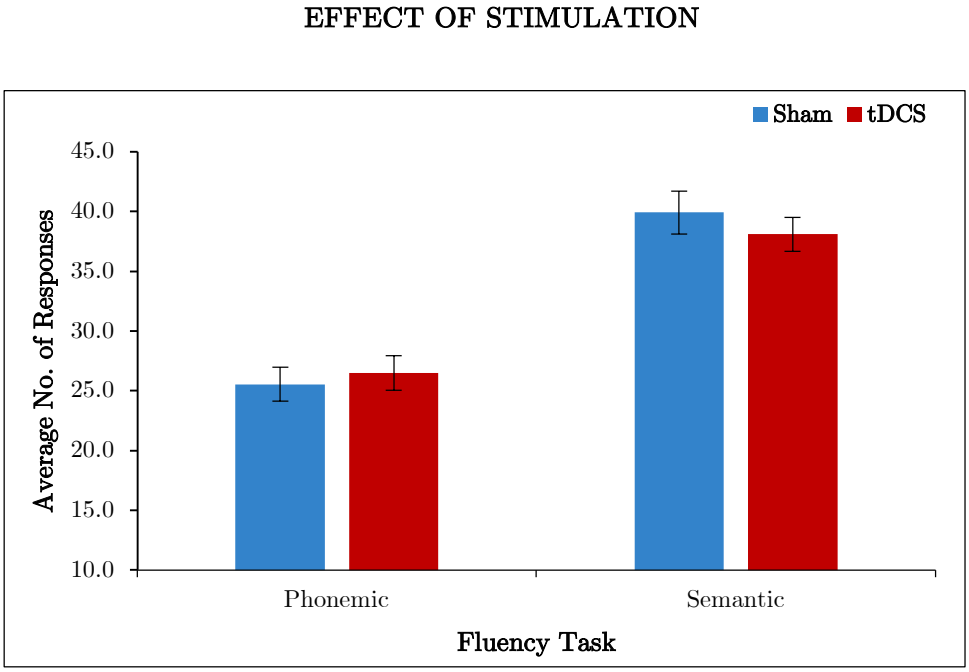


Fig 5.1 – Average number of responses across tasks and stimulation conditions. Error bars reflect Standard Error

5.4 Experiment 2: Meta-analysis of tDCS effects on verbal fluency

Here, I conducted a systematic review and meta-analysis to re-evaluate the effects of anodal tDCS when applied in a single session to healthy participants. I generate effect size estimates for the overall effect of tDCS on fluency tasks, but also break down effects based on phonemic and semantic fluency performance. I further provide separate estimates with all current published data and all published data plus the replication presented above.

5.4.1 Method

I provide only necessary information since the method is a direct replication of previous work reported in Chapter 3.

5.4.1.1 Literature Search

I searched Science Direct, Web of Knowledge and PubMed databases (from 1999 to early August 2017) using as search keywords: 'tDCS' or 'transcranial direct current stimulation' in combination with, 'fluency', 'semantic', 'phonemic', 'language', 'word production', 'speech fluency', and 'cognition'. Web of Knowledge tracking tool to extract papers in references and citations. The initial search returned 3,960 articles of which 3,411 were removed right away as non-relevant. The text of the remaining 549 papers were read (papers testing neurologically impaired individuals were also searched for healthy control samples), of which 534 studies were excluded because fluency and item recognition tasks were not measured in healthy participants, leaving 11 articles. One paper was excluded because it investigated the effects on individual who stutter (Chesters, Watkins & Möttönen, 2017). This left 10 papers, with a total number of 18 conditions (within-samples, $n = 10$, between-subjects, $n = 8$) in which tDCS was measured (see Appendix 5.1).

5.4.1.2 Data Extraction

We focused only on conditions using anodal tDCS applied to left frontal regions. Thus, means and standard deviations for average number of responses for anodal tDCS and sham conditions were extracted. Plotted values were converted into numerical values using a java program called Plot Digitzer (Joseph, 2011; for similar method, see Hill et al., 2015; Westwood & Romani, 2017; Vashegi et al., 2015). Data not reported was requested by email.

In one instance (Cerruti & Schlaug, 2009, Exp 2), only the t-value was reported, which we converted into an effect size estimate using the following formulae:

$$\text{Cohen's } d = t \times \sqrt{\left((n_1 + n_2)/(n_1 \times n_2)\right)}$$

5.4.1.3 Data Analysis

Direction of the effect. The effect of tDCS was quantified by expressing the difference between tDCS and sham conditions as the number of standard deviations – i.e., the difference between conditions over a measure of variability. Our hypothesis was that anodal tDCS of the left frontal regions would enhance verbal fluency. Effect size estimates coded in the positive direction are consistent with this prediction, negative otherwise.

Effects from within- and between-participant studies. Previous reviews typically calculate effect sizes assuming a between-participants design (see Horvarth et al., 2015; Jacobson et al., 2012; Price et al., 2015), but this overestimates variance if the design was within-participants (because some variance is shared across conditions), thereby reducing the likelihood of finding significant result. In this meta-analysis, I used different effect size estimates for within- and between-study designs to increase precision (see Borenstein et al., 2009; Lakens, 2013).

Multiple dependent effects. In a meta-analysis, one assume the summary effect size is based on effect sizes drawn from different participant samples. However, several studies reported effects of tDCS on phonemic and semantic fluency performance, so if one were to include these separately in one meta-analyses it would lead to an underestimation of variance and an overestimation of statistical significance – i.e., a Type 1 Error (Lipsey & Wilson, 2001). Convention is to select one effect per study, but this reduces power (see Jacobson et al., 2012; Price et al., 2015). I used composite effects across phonemic and semantic conditions, calculated as the mean performance and variance that also considered the correlation of variance (or error) between conditions (for method, Chapter 3; also, see Borenstein et al.,

2009). Unfortunately, in the case of Pisoni et al. (2017), only mean performance across semantic and phonemic fluency was reported – I was unable to obtain data separately.

Heterogeneity. This was measured using *Cochran's Q statistic* (see Borenstein et al., 2009). A significant Q indicates significant heterogeneity, but because Q suffers from low power with small sample sizes, it is difficult to equate a non-significant Q with no heterogeneity. We therefore increased the threshold p value to .10 (Higgins et al., 2003), and quantified heterogeneity as a percentage using the I^2 index. A rule of thumb is that I^2 indexes of 75%, 50%, and 25% reflect high, medium and low true heterogeneity, respectively (Lipsey & Wilson, 2001).

Fixed vs. Random Effects. A fixed effects model assumes variation in effect size across studies is only due to sampling error. A random effects model, however, assumes the variation across studies is *due to a combination of sampling error and* differences in methodology, which is important since tDCS studies vary in term of applied stimulation parameters. We therefore used a random effects model.

Outliers. We planned to exclude effect size estimates 3*SD* above or below the summary effect size we excluded, but in all analyses, no study met this criterion.

5.4.2 Results

5.4.2.1 Primary Analysis – Overall effect of tDCS

Figure 5.1 presents funnel plots for effects across all pooled studies. Effects are reported as positive if consistent with the general hypothesis that left anodal tDCS improves performance; negative otherwise. The results showed a moderate but significant effect size (Hedges' g = .36, 95%CI (0.07, 0.65)) associated with significant heterogeneity between effect size estimates (Q = 48, df = 10, p < .001, I2 = 79%).

5.4.2.1 Secondary Analysis – Effect of tDCS by Fluency Task

Figure 5.2 presents funnel plots for effects across studies pooled according to type of fluency of task. Unfortunately, Pisoni et al. (2017) was excluded from this analysis because they only reported the effect of tDCS averaged across phonemic and semantic fluency tasks, we

were unable to obtain data separately for each task. The results showed a null effect size on phonemic fluency (Hedges' g = .01, 95%CI (-0.14, 0.15)), with no evidence of publication bias, but I did find a positive effect on semantic fluency (Hedges' g = .52, 95%CI (0.14, 0.89) with evidence of publication bias. However, I interpret the finding of publication bias with caution. This is because the analysis of publication bias may have been confounded by the significant heterogeneity seen between effect size estimates thanks to several larger than typical effects (see Cattaneo et al., 2011; Martin et al., 2017; Meinzer et al., 2012) which would have exaggerated the asymmetry between positive and negative effects (as can be seen in Figure 5.2c).

PRIMARY ANALYSIS – Overall Effect of Anodal tDCS

Study	Hedges' *g*	Lower	Upper
Cattaneo et al. (2011)	0.90	0.28	1.52
Cerruti & Schlaug (2008), Exp 1	-0.15	-0.63	0.32
Cerruti & Schlaug (2008), Exp 2	0.14	-0.33	0.62
Ehlis et al. (2016)	0.06	-0.30	0.41
Martin et al. (2017)	1.50	0.99	2.02
Meinzer et al. (2012)	0.69	0.27	1.12
Penolazzi et al. (2013)	0.20	-0.43	0.83
Pisoni et al. (2017)	0.70	0.26	1.15
Vannorsdall et al. (2012)	0.20	-0.28	0.67
Vannorsdall et al. (2016)	0.01	-0.43	0.45
Westwood & Romani (submitted)	-0.10	-0.34	0.15
Average	**0.36**	**0.07**	**0.65**
Heterogeneity	$Q = 48$, df $= 10$, $p < .001$, $I^2 = 79\%$		
My replication of Cattaneo et al (2011)	-0.06	-0.40	0.29
Overall Average	**0.32**	**0.05**	**0.58**
Heterogeneity	$Q = 51$, df $= 11$, $p < .001$, $I^2 = 78\%$		

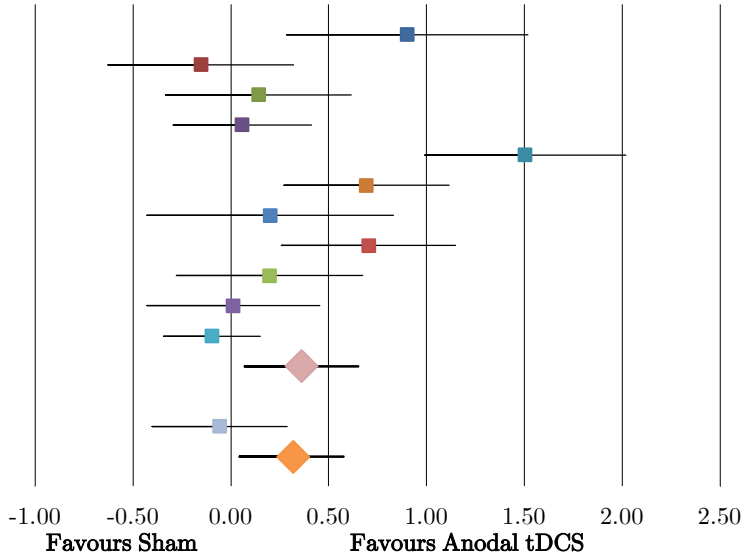


Fig 5.2 – Forest Plots for the size of tDCS effects on average number of response across fluency tasks. Error bars reflect 95% confidence intervals. Effects size given in Hedge's g.

SECONDARY ANALYSIS – Effect of tDCS on Phonemic Fluency Task

Study	Hedges' *g*	Lower	Upper
Cattaneo et al. (2011)	0.50	-0.05	1.04
Cerruti & Schlaug (2008), Exp 1	-0.15	-0.63	0.32
Cerruti & Schlaug (2008), Exp 2	0.14	-0.33	0.62
Ehlis et al. (2016)	-0.05	-0.40	0.31
Vannorsdall et al. (2012)	-0.22	-0.70	0.25
Vannorsdall et al. (2016)	0.19	-0.25	0.64
Westwood & Romani (2017a)	-0.06	-0.30	0.19
Average	**0.01**	**-0.14**	**0.15**
Heterogeneity	$Q = 6$, df = 6, $p = .45$, I2 = 0%		
My replication of Cattaneo et al (2011)	0.14	-0.21	0.49
Overall Average	**0.03**	**-0.11**	**0.16**
Heterogeneity	$Q = 6$, df = 7, $p = .51$, I2 = 0%		

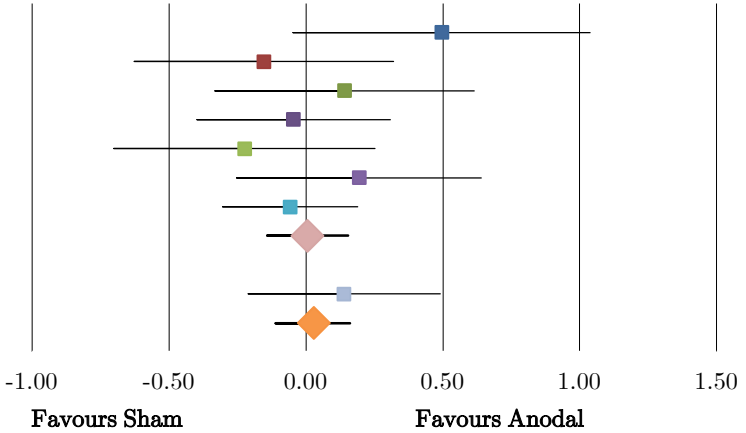


Fig 5.3 – Forest Plots for the size of tDCS effects on average number of response across phonemic fluency tasks. Error bars reflect 95% confidence intervals. Effects size given in Hedge's g.

SECONDARY ANALYSIS – Effect of tDCS on Semantic Fluency Task

Study	Hedges' *g*	Lower	Upper
Cattaneo et al. (2011)	1.07	0.41	1.73
Ehlis et al. (2016)	0.12	-0.24	0.47
Martin et al. (2017)	1.50	0.99	2.02
Meinzer et al. (2012)	0.69	0.27	1.12
Penolazzi et al. (2013)	0.20	-0.43	0.82
Vannorsdall et al. (2012)	0.45	-0.04	0.95
Vannorsdall et al. (2016)	0.43	-0.04	0.89
Westwood & Romani (2017a)	-0.11	-0.36	0.14
Average	**0.52**	**0.14**	**0.89**
Heterogeneity	$Q = 41$, df = 7, $p < .001$, I2 = 79%		
My replication of Cattaneo et al (2011)	-0.21	-0.57	0.14
Overall Average	**0.43**	**0.08**	**0.78**
Heterogeneity	$Q = 50$, df = 8, $p < .001$, I2 = 84%		

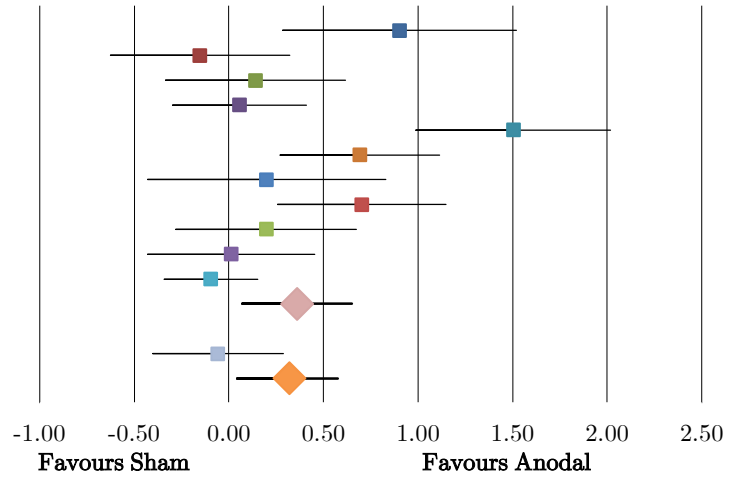
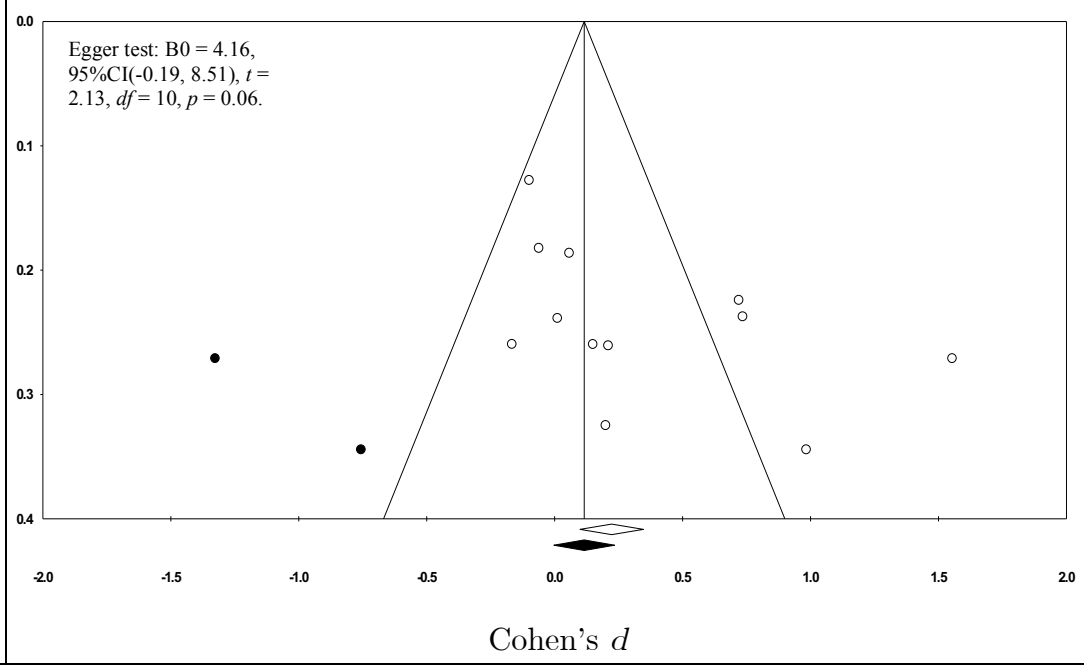


Fig 5.4 – Forest Plots for the size of tDCS effects on average number of response across semantic fluency tasks. Error bars reflect 95% confidence intervals. Effects size given in Hedge's g.

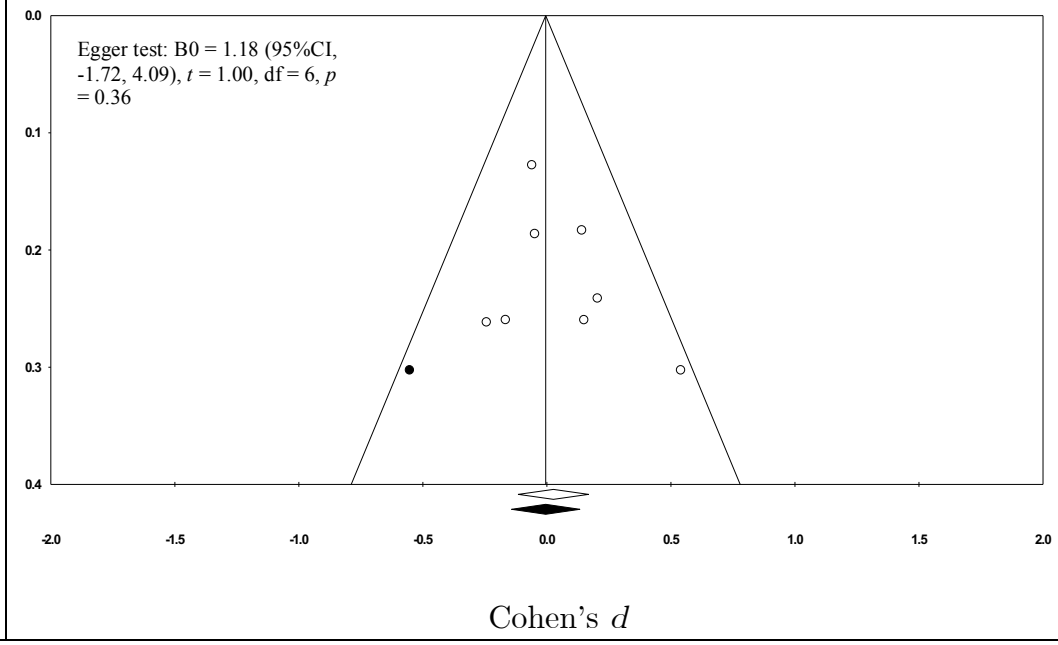
a. Overall Performance



Trim and Fill

Pre: 0.33 (95%CI 0.06, 0.61); Post: 0.16 (95%CI -0.16, 0.48)

b. Phonemic Fluency



Trim and Fill

Pre: 0.03 (95%CI, -0.12, 0.17); Post: -0.00 (95%CI, -0.16, 0.16)

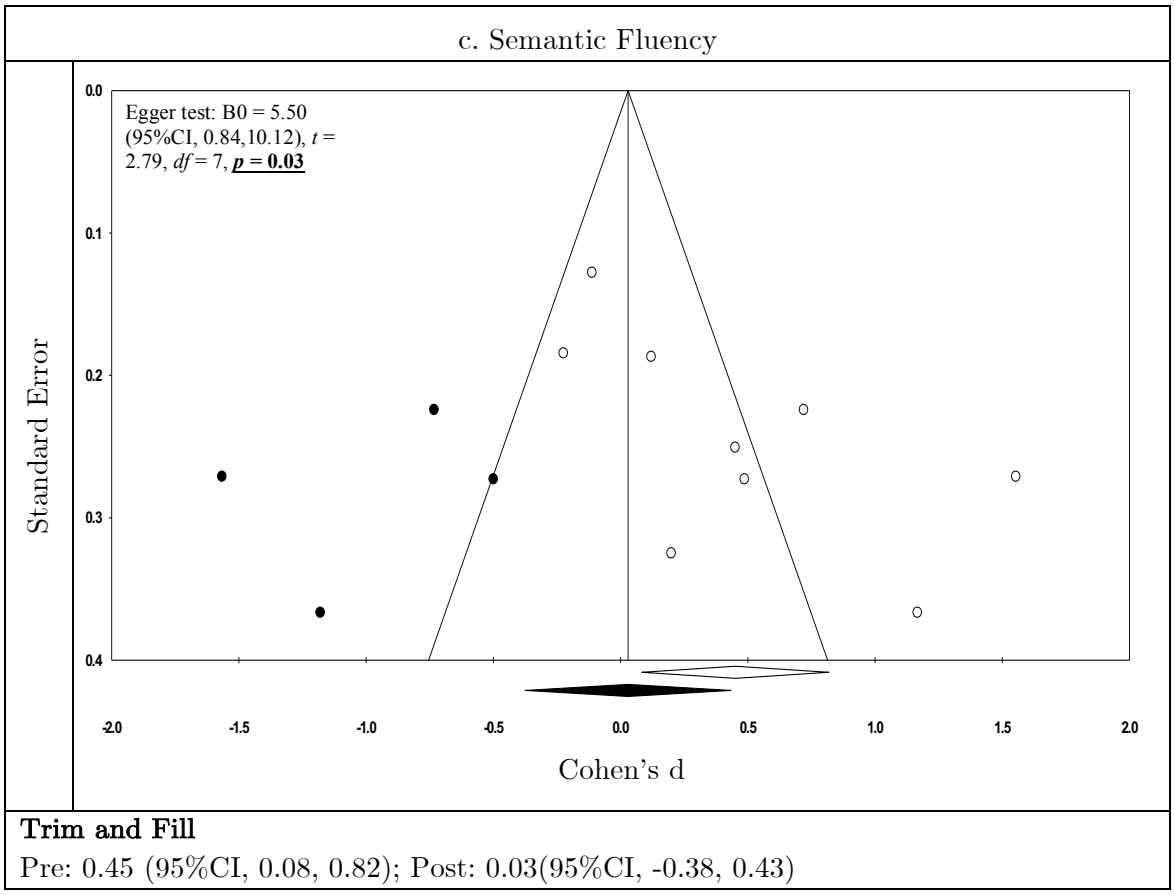


Fig 5.4 – Funnel plots when pooling studies for overall performance and for phonemic and semantic fluency tasks with my replication of Cattaneo et al. (2011). Effect size estimates provided before and after trim and fill procedure. Effects size given in Cohen's d.

5.5 Discussion

Inconsistent effects are reported across studies measuring the effects of single session tDCS on verbal fluency tasks in healthy participants. Attempts to address this problem in the form of quantitative reviews or meta-analysis lead to conflicting results, likely due to methodological differences in pooled studies (Horvath et al., 2015; Price et al., 2015), but also from large treatment effects typical in fields with underpowered studies and publication bias. I investigated this latter explanation by carrying out: 1) a direct replication of a previous study; and 2) a meta-analysis that pooled data from studies measuring the effects of anodal tDCS when applied to the left frontal region in healthy participants, including my replication. The meta-analysis showed a positive effect of anodal tDCS on verbal fluency, with semantic fluency tasks being more amenable to the effect of tDCS compared to phonemic fluency. However, for this last finding, I also found significant publication bias in the expected direction – i.e., in line with the expectation that anodal tDCS should improve performance. Yet, this finding is better explained with evidence of heterogeneity between effect size estimates, which would exaggerate differences between positive versus negative effect sizes, and – as suspected in the introduction – instead suggested that the significant effect of tDCS likely emerged from exceptionally large effect sizes reported by a minority of studies (i.e., Cattaneo et al., 2011; Martin et al. 2017; Meinzer et al., 2012; Pisoni et al., 2017). I attempted to replicate one of these studies, but the results showed no effect of tDCS. Moreover, when this replication was included in the meta-analysis, the summary effect size was reduced, suggesting that these exceptionally large effects are likely a chance result caused by natural variability in the effect of tDCS. This is important to consider given the impact these effects had on the meta-analysis outcome, a tool that is frequently used to draw inferences about the efficacy of tDCS.

The exceptionally large effect estimates may of course be because they came from studies using a particularly effective combination of stimulation parameters. However, I find no evidence of this. I have failed to reproduce the results seen by Cattaneo et al (2011; see also Penolazzi et al., 2013a; Vannorsdall et al., 2016), and there does not seem to be a consistent association between a particular protocol and positive results (see Appendix 5.1). For instance, Cattaneo et al. (2011) applied 2mA anodal tDCS ($0.06mA/cm^2$; $35cm^2$ active

electrode) for 20mins to the LIFG before verbal fluency performance; Martin et al. (2017) applied 1mA anodal tDCS (or 0.03mA/cm^2; 35cm^2 active electrode) for 30mins to the primary motor cortex during semantic fluency; Meinzer et al (2012) applied 1mA anodal tDCS (or 0.03mA/cm^2; 35cm^2 active electrode) for 17mins to the LIFG plus anterior temporal regions during semantic fluency performance; and, finally, Pisoni et al. (2017) applied .75mA of anodal tDCS or (0.05mA/cm^2; 16cm^2 active electrode) for 20mins to the LIFG during verbal fluency performance via smaller rounded scalp electrodes.

Alternatively, it may be that large effects emerged because of the type of task. Martin et al. (2017) and Meinzer et al. (2012) both used variants of the typical semantic fluency task in which participants gave ten examples for six different semantic categories, and the task was also self-paced. It may be that larger effects are more likely in this variant of semantic fluency, or semantic fluency in general given that the results showed semantic fluency to be more amenable to the tDCS effect. However, there is no theoretically plausible justification for this, especially since a majority null effect size estimates were based on data from studies using the typical semantic fluency task (i.e., produce as many examples in response to a category cue within a given time limit of sixty seconds), whilst the other two studies that reported large effects (i.e., Cattaneo et al., 2011 and Pisoni et al., 2017) did not use the variant of semantic fluency used by Martin et al (2017) and Meinzer et al. (2012).

My failed replication may have arisen from factors other than the poor efficacy of tDCS. The replication differed from Cattaneo et al. in terms of sample size, sham protocol, and stimuli. However, the increase in sample size should have in fact improved the chances of obtaining a significant effect, and the sham protocol has been proven to be effective in this and previous studies (see Westwood et al., 2017). In regards the stimuli, as Cattaneo et al. (2016) pointed out, differences in stimuli can be tolerated in replications, so long as stimuli are not repeated across sessions, to avoid practice effects. I used different stimuli in each session, and counterbalanced the order by session and stimulation condition across participants. This is in keeping with the recommendation made by Cattaneo et al. (2016).

Another potential factor was lack of power to find a significant outcome. A power calculation showed that my sample size of 24 participants was sufficiently large enough to detect an

effect of the size reported by Cattaneo et al. (˜1.1, 1- β = .97). However, this effect arose from a very small sample size, and is therefore likely to be an exaggeration of the true effect. In my meta-analysis, comprised of roughly 230 participants, I estimate the size of the effect to be .32. To detect an effect this size, my study and Cattaneo et al.'s 2011 study were massively underpowered (1- β = .38 versus .20). It is, therefore, possible that the effect reported in Cattaneo et al was in fact a chance result that overestimated the true effect of tDCS, since underpowered studies are more likely to detect large effects (see Button et al., 2013). The same can be said for the large effects reported by Martin et al. (2017), Meinzer et al. (2012), and Pisoni et al. (2017), all of which recruited fewer than 24 participants. Thus, future replications with increased sample sizes should be carried out for more precise estimates of the effect of tDCS (see also Medina & Cason, 2017).

Chapter 6: Investigating the effects of multi session tDCS on word learning in healthy participants

6.1 Overview

In the previous chapters, I demonstrated that effects of tDCS are unreliable when applied in a single session to healthy participants during word production and working memory tasks. In this chapter, I describe preliminary work investigating the effects of anodal tDCS when applied to the LIFG over multiple sessions during a language learning paradigm. Several lines of research suggest that LIFG plays an important role in learning by selecting unstable target information from long-term memory. This is consistent with research demonstrating its involvement in retrieval of lexical semantic representations from long-term memory (Buckner et al., 2000; Gabrieli et al., 1996; Petersen et al., 1988; Price, 2000; Thompson-Schill et al., 1999; Wise et al., 1991), and studies that show functional connectivity to the left posterior middle temporal gyrus (considered the neural locus of the mental lexicon, Buckner et al., 2000; Thompson-Schill et al., 1999). Thus, I outline a plan for future experiments to assess the effectiveness of alternative conditions that may increase the likelihood of a significant effect when targeting the LIFG to enhance novel language learning.

6.2 Introduction

The unreliable effects of tDCS seen across the literature and the null effects I report are clearly not an encouraging indication of its effectiveness. However, this does not mean that we should scrap this technique completely. Much is still to be learned about the conditions in which it is and is not effective. Null effects may have emerged because there is little scope to enhance cognitive abilities with single sessions of tDCS in healthy participants. In the healthy brain, cortical excitability is operating already at or near to optimal levels, and homeostatic mechanisms – which dynamically adjust excitability to promote stability of neural networks – may counter the weak effect tDCS has on cortical excitability (see Amadi, Allman, Johansen-Berg & Stagg, 2015; Besson, Perrey, Teo & Muthalib, 2016; Giordano et al., 2017; Hoy et al., 2013). The implication follows that reliable and/or detectable changes

in cognition are more likely when neural resources are not operating at optimal levels (for similar arguments, see Giordano et al., 2017; Hill et al., 2015; Hoy et al., 2013). Thus, I want to build on my research to disentangle two possible, but not mutually exclusive, variables in explaining differences in the efficacy of anodal tDCS. These are: type of population and type of protocol.

Positive effects of anodal tDCS maybe more likely in neurologically damaged individuals where there is pathologically low levels of excitability and dysfunctional homeostatic mechanisms (Murphy & Corbett, 2009). Compared to research on healthy participants, picture naming, word reading, and word re-learning abilities have been more systematically improved following courses of tDCS in stroke induced aphasic patients (for reviews, see Crinion, 2016; de Aguiar et al., 2015). There is also evidence of positive cognitive outcomes in patients with Parkinson's (Lawrence, Gasson, Bucks, Troeung, & Loftus, 2017), Schizophrenia (Pondé et al., 2017), Alzheimer's (Zhao et al., 2017) and dipolar depression (Donde et al., 2017; see also Cappon et al., 2016; Zhao et al., 2017). Consistent with this literature is evidence of positive effects observed in healthy older adults (but see Fertonani et al., 2014), where there is age-related decline in neuronal functioning (Bishop, Lu & Tanker, 2010). Ross et al. (2011) found that anodal tDCS applied to the left or right anterior temporal lobe facilitated naming accuracy particularly for harder to name items, with the magnitude of facilitation being smaller and limited to right hemisphere stimulation in younger adults. Berryhill and Jones (2012), on the other hand, found an anodal tDCS related improvement across visual and verbal working memory tasks in older adults with higher education levels.

Alternatively, effects may be seen even in healthy participants, but only during learning, such as during novel language acquisition. In the course of learning, new representations become established following a period of instability (Takeuchi, Duszkiewicz & Morris, 2014), and it is in this period – characterised by low neuronal functioning – where the effects of tDCS may preferentially operate. This scenario contrasts with my previous experiments, in which tDCS was operating on representations that have been well consolidated over many years across a distributed network of cortical regions (see Jacobson et al., 2012; Reinhart et al., 2017). That tDCS can modulate learning is in line with evidence of tDCS related changes

in neurotransmitter and neuromodulator activity (e.g., glutamate, GABA, dopamine, serotonin, and acetylcholine), both of which underpin synaptic plasticity (see Medeiros et al., 2012; Stagg & Nitsche, 2011). Moreover, studies recruiting healthy participants have shown enhanced learning of novel languages (Fiori et al., 2011; Flöel et al., 2008; Liuzzi et al., 2010), artificial grammar (DeVries et al., 2009), novel face-name associations (Matzen, Trumbo, Leach & Leshikar, 2015; Pisoni, Vernice, Iasevoli, Cattaneo & Papagno, 2015), and memorising lists of words (Elmer, Burkard, Renz, Meyer & Jancke, 2009; Javadi & Cheng, 2013; Javadi, Cheng & Walsh, 2012; Javadi & Walsh, 2012).

Additionally, positive effects in healthy participants are more likely when novel word learning paradigms are coupled with tDCS applied over multiple sessions. Studies using healthy participants show that neurophysiological effects of tDCS can accumulate if administered over several sessions (Fricke et al., 2011; Monte-Silva et al., 2013; Bastani & Jaberzadeh, 2014), with greater gains observed in terms of smoking craving (Boggio et al., 2009), food craving (Ljubisavljevic, Maxood, Bjekic, Oommen & Nagelkerke, 2016), reading (Heth & Lavidor, 2015), and motor learning (Gallasch, Christova, Rafolt & Gallasch, 2015; Reis et al., 2009). Consistent with this evidence is the fact that studies with aphasic patients usually administered tDCS in conjunction with naming across multiple sessions, raising the possibility that positive effects seen in this population resulted in part from the repeated exposure to tDCS. Importantly, positive effects with healthy participants have been reported with repeated tDCS stimulation in conjunction with learning an artificial number alphabet (Cohen Kadosh et al., 2010), and learning a working memory task (Talsma, Kroese & Slagter, 2017), with one study detecting improvements in novel word learning after five daily-sessions of anodal tDCS (Meinzer et al., 2014).

6.3 Present experiment: alternate day tDCS and word learning

Here, I wanted to measure the effectiveness of tDCS when applied in multiple sessions during a word learning paradigm. Participants were asked to learn sixty Italian words paired with pictures over three sessions (Monday, Wednesday, and Friday) – the modality of learning was chosen because it was similar to what would be used to learn the vocabulary of a foreign language. I also assessed the consolidation of learned words in subsequent sessions, and

carried out a follow-up assessment one week later. In each session, I applied anodal tDCS to the LIFG, with no stimulation applied in the final follow-up session. I chose a between-participants design, whereby in one group word learning was paired either with tDCS or sham stimulation. The hypothesis was that anodal tDCS would significantly improve language acquisition, with effects accumulating across sessions, and maintained one week later. This is in line with previous studies which show a) involvement of the LIFG in lexical access and learning (Price, 2012); b) anodal tDCS applied to the LIFG can enhance learning (DeVries et al., 2009; Meinzer et al., 2014); and c) that effects accumulate over repeated sessions mentioned previously. In anticipation of my findings, I was able to detect a beneficial effect of tDCS, particularly in terms of the rate learning. However, these results are preliminary. The plan is to collect data from sixty participants (thirty in each stimulation condition), but so far I have collected data from twenty participants (ten in each stimulation condition). The small sample size, extreme variability in learning skills across participants, and the variation introduced by the between-participants design, suggests that more participants will be needed to have enough power to detect positive effects of tDCS.

6.3.1 Method

6.3.1.1 Procedure Overview

The experiment was conducted over four testing sessions. Across the first three (so-called *learning)* sesssions, participants were asked to learn 60 Italian words for everyday objects (e.g. *ponte* for 'bridge'), with 20 words being learned in each session. A fourth (so-called *maintenance*) session was conducted as a follow-up to test the long-term renention for all words. The learning sessions each took 30 to 50mins to complete and were conducted on alternate days of the same week (i.e., Monday, Wednesday, Friday), whilst the maintenance session took roughly 20 to 30mins to complete and was conducted on the following Thursday. Participants received either sham or real tDCS, and allocation to stimulation conditions was completely random. The real tDCS group were administered anodal tDCS in the learning sessions for up to 25mins, covering the entirety of the encoding, acquisition, testing and consolidation phases of the word learning task (more details below). An experimenter was present throughout the task performance (sat directly behind and to the right of

participants) to record accuracy of responses. Stimuli were presented via a laptop computer, which also recorded response times (see Stimuli section below for details).

6.3.1.2 Design

Learning Sessions (~15 to 20mins). Each of the three learning sessions involved three main phases, encoding, acquisition, and testing. In addition, to consolidate learning, a *consolidation* phase was conducted in the second and third session in which words presented in previous sessions were presented again at the beginning of the session, *before* the three main phases. Stimulation was applied at the beginning of each learning session, thereby covering all learning and consolidation phases (see Figure 6.1 and Table 6.1).

Encoding phase (picture presented with words; ~ 5mins). In this phase, 20 pictures were presented one after the other with the corresponding Italian and English word written below and an auditory recording of the correct pronunciation of the Italian word. Participants were asked to repeat the Italian word for each presentation, and memorise its association with the pucture. Participants were asked to go through the stimuli at their own pace. The entire set of picture-word pairs were presented in 2 cycles, with an optional break provided after the first presentation. Participants took about 4 to 5mins to complete this phase, and no participant took the optional break.

Acquisition phase (pictures only, with feedback; ~7mins). Once the Encoding phase was finished, the Acquisition phase began. In this phase, pictures were presented without any paired word (visual or auditory), and participants were asked to name the pictures with the correct corresponding Italian word. The experimenter provided feedback on each trial if the participant said the word incorrectly or said an alternative Italian word, in which case the experimenter provided the correct pronunciation and the participant was asked to repeat it. Pictures were presented over 8 cycles, with a short optional break provided between each cycle. Participants took no more or less that 6 to 7 mins to complete this phase, and no participant took the optional breaks. Accuracy of responses was recorded during the session by the experimenter.

Testing phase (pictures only, without feedback; ~2mins). Once the Acquisition phase was finished, the Testing phase. In this phase, the same pictures were presented again without any word or feedback on response accuracy. Pictures were presented over 2 cycles, with no break between cycles. Participants were asked to respond as accurately as fast as possible, and were reminded that no feedback would be given. Participants took about 1 to 2 mins to complete this phase

Additional Consolidation Phase (~5mins). This phase was only applicable to the second and third session, and was designed to consolidate and to test the memory of words learned from previous sessions. At the beginning of the second and third session, before participants started the three main phases (mentioned directly above), previously presented sets of pictures were presented in two phases, *re-acquisition* and *re-testing*. In the *re-acquisition* phase, participants were asked to name pictures with feedback provided by the experimenter. In the *re-testing* phase, participants were asked to name as fast and as accurately as possible the same pictures. In both phases, the picture sets were presented repeatedly over several cycles. The number of cycles differed according to session, which ensured that participants did not over learn stimuli and to ensure that all task performance fell with the duration of stimulation (see Table 6.1).

Maintenance Session (~20mins). In this phase, participants were asked to naming pictures without feedback, with feedback, and then again without feedback. Participants named pictures over two cycles, with optional breaks given between each cycle and between each naming blocks.

6.3.1.3 Stimuli

We selected 60 Italian bisyllabic nouns from a corpus of 626 Italian nouns (Barca, Burani, & Arduino, 2002; see Appendix 6.1). Words were carefully selected to avoid, a) similarities to the English word (e.g., *Barba* for 'beard'); b) cognates of English words (e.g., *sede* for 'seat'); c) similarities to other English words (so-called false friends; e.g., *calamita* for 'magnet'); and d) speech sounds not found in the English language (e.g., *coniglio* for 'rabbit'). The words were divided into 3 sets of 20 words (sets A, B and C). Words were carefully matched across sets in terms of typical age of acquisition, frequency, word length

and consonant-vowel structure, and syllabic structure to match for difficulty. 60 coloured pictures that corresponded to Italian words were sourced from Google images, which were judged by three experimenters as unambiguous depictions of each Italian word (See Appendix 6.1).

Stimuli were presented using E-Prime 2 Software and a Dell Laptop computer screen (screen size: 15.6"). Pictures were presented centre of the screen (720 x 540 pixels), words (Italian and English) were presented in Arial typeface 24-font directly underneath pictures. Each presentation trial began with a fixation cross presented in the centre of the screen for 1000ms, followed by a blank screen for 250 msec, then the stimuli was presented (pictures/words/audio) until the participant pressed a key or mouse button, and finally a blank screen 500 msec appeared before the next trial began. For the encoding phase, picture-word pairs were presented with auditory recording; for the acquisition and re-acquisition phase, only pictures were presented. Because the encoding, acquisition and re-acquisition phases were designed to be self-paced, the stimuli stayed on the screen until participants pressed any button on the keyboard to continue to the next trial. In the testing and re-testing phases, only pictures were presented. These were presented for 2500ms or until a response was picked up from the voice key, after which the next trial was initiated after a blank screen presented for 500ms. Vocal responses were recorded using a Sony ICDPX333.CE7 voice recorder. The voice key was a serial response box (Refresher Detector System, Psychology Software Tools, INC). The microphone was a Sony ECM-MS957.

6.3.1.4 transcranial direct current stimulation (tDCS)

Stimulation was administered via a battery driven NeurConn DC-Stimulator using 25cm^2 ("active") and 35cm^2 ("reference") electrode sponges soaked in saline solution. The active electrode was placed on the LIFG, whilst the reference electrode was placed on the contralateral supraorbital area. The LIFG was located as F7 in the 10/20 EEG system, which we located by measuring 2cm from the corner of the eye to the ear then 3cm at perpendicular upwards (see Devlin & Watkins, 2007). For sham stimulation, an intermittent current of 110 mA was delivered for a period of 3 msec every 550 msec. This produces the perceptual sensations of real stimulation without modulating underlying brain areas (Palm

et al., 2013). For anodal tDCS, we administered a 1.5 mA current for 25mins in conjunction with learning and consolidation phases of the word learning tasks (see Figure 6.1).

EXPERIMENT PROCEDURE TIMLINE

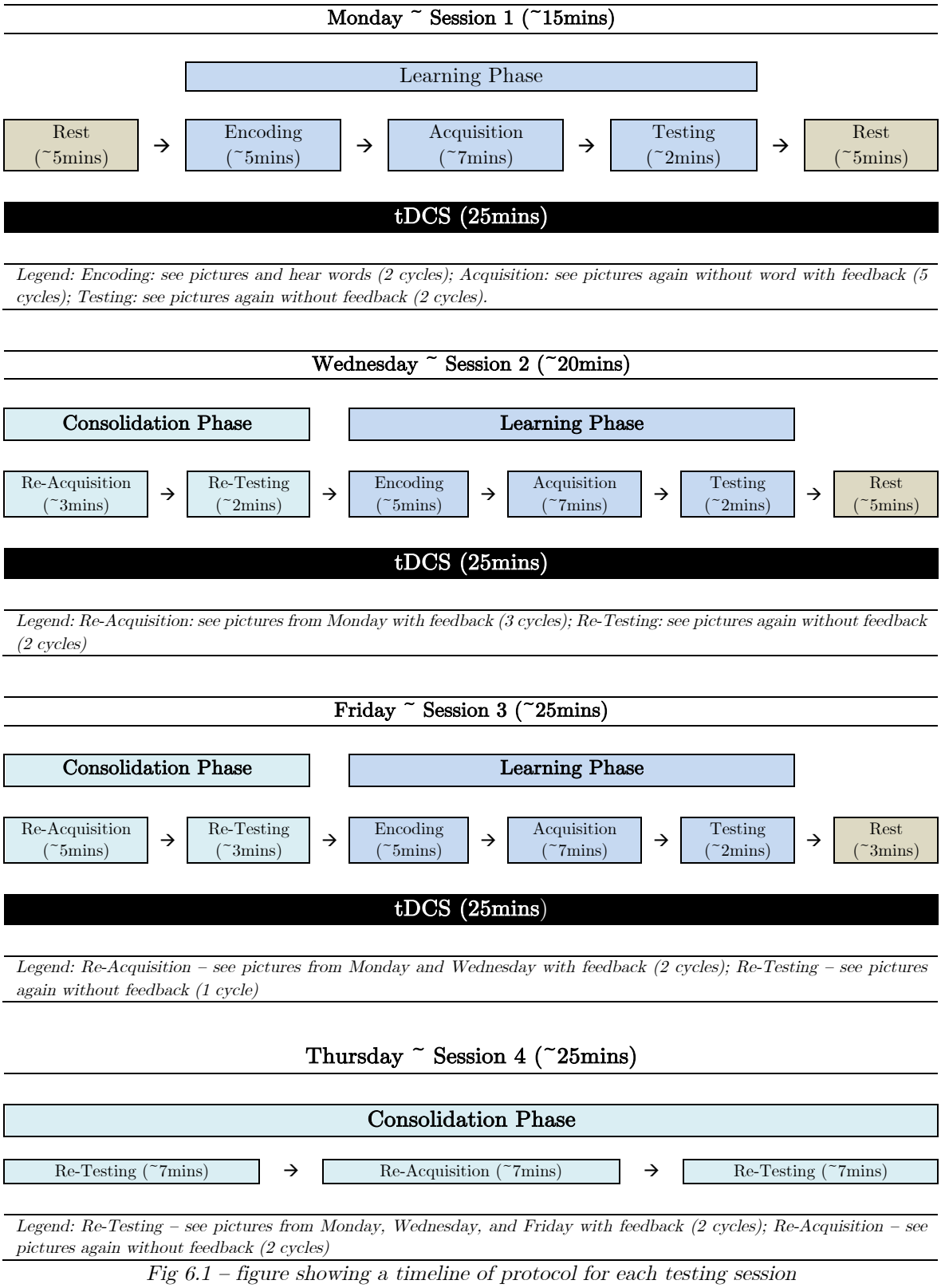


Fig 6.1 – figure showing a timeline of protocol for each testing session

Table 6.2 – table showing the number of cycles for a given set presented on a given day

Stimuli Set	Session				Total Cycles
	Monday	Wednesday	Friday	Thursday	
Monday	9	5	3	6	23
Wednesday	N/A	9	3	6	18
Friday	N/A	N/A	9	6	15

6.4.2 Results

6.4.2.1 Encoding Phase

Figure 6.2 shows the average time participants took to complete the encoding phase across each testing session and stimulation conditions. It is clear that neither sham nor real tDCS impacted on the time taken to encode stimuli. This was confirmed in a mixed factors ANOVA, with *Condition* (Sham versus tDCS) as a between participants factor, *Session* (Monday, Wednesday or Friday), and *Cycle* (1 versus 2) as a within participants factor, and time taken to complete encoding phases in each session as the dependent measure. The results revealed no significant main effect of *Condition* ($F(1,18) = .04$, $p = .84$, $\eta_p^2 = .002$), *Session* ($F(2,36) = 2.69$, $p = .08$, $\eta_p^2 = .13$), but there was a borderline main effect of *Cycle* ($F(2,36) = 4.37$, $p = .05$, $\eta_p^2 = .20$), with participant taking longer to go through stimuli in the first compared to the second cycle (8.2±.89 versus 6.3±.59 seconds), which is an expected effect of practice. Crucially, there was no significant interactions for *Condition* x *Session* interaction ($F(2,36) = .04$, $p = .96$, $\eta_p^2 = .002$), *Condition* x *Cycle* ($F(1,18) = 1.59$, $p = .22$, $\eta_p^2 = .08$), *Session* x *Cycle* ($F(2,36) = .20$, $p = .82$, $\eta_p^2 = .01$), *Condition* x *Session* x *Cycle* ($F(2,36) = .11$, $p = .90$, $\eta_p^2 = .01$).

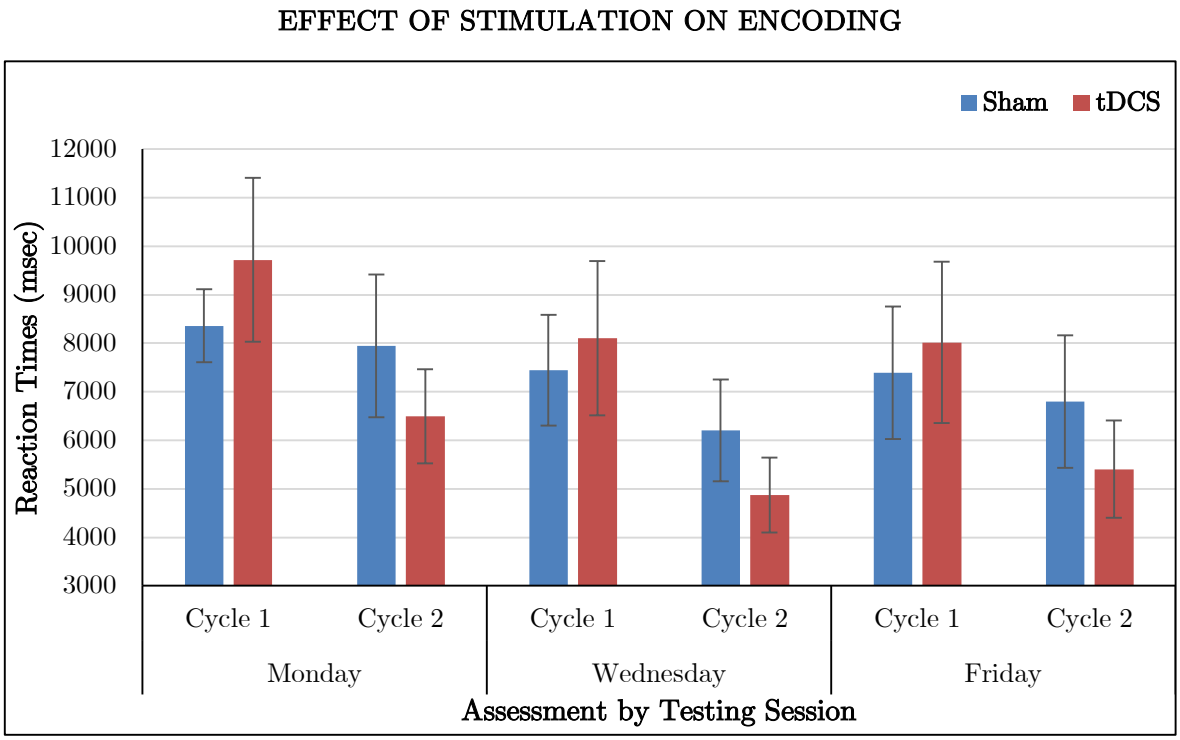


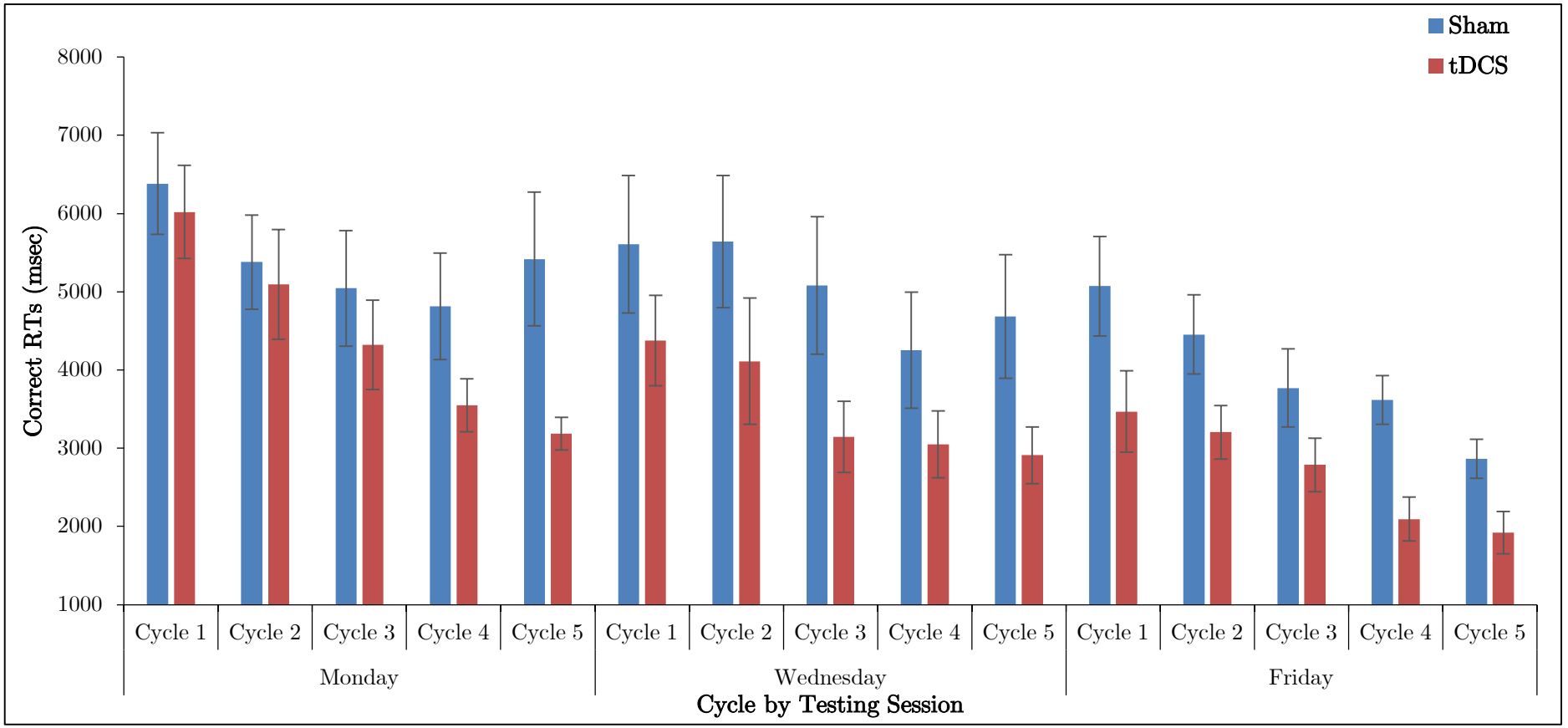
Fig 6.2 – average time taken (msec) during the encoding phase across stimulation conditions, cycles, and testing session. Error bars reflect standard error.

6.4.2.2 Acquisition Phase

Here, I investigated the rate of learning in terms of the average percentage of errors and reaction times when responding to stimuli (with feedback) across the five cycles within and across each session. Figure 6.4.2 and Figure 6.4.3 shows that in both stimulation groups, percentage of errors and reaction times reduced with each cycle. Moreover, the magnitude of this reduction was also a function of session, with performance generally improving on Wednesday compared to Monday. This is most likely due to practice effects, which is to be expected as participants became used to the task. Participants, however, were markedly less error prone and faster in tDCS compared to sham across cycles, with this difference increasing with each testing session.

I carried out separate mixed factors ANOVAs on percentage errors and reaction times, with *Condition* (Sham versus tDCS) as a between-participants factor, *Session* (Monday, Wednesday or Friday) and *Cycle* (1 to 5) as a within participants factor. The results showed a significant main effect of *Condition* ($_{RTs}F(1,18) = 5.36$, $p = .03$, $\eta_p^2 = .23$), with participants being markedly faster when given anodal tDCS compared to sham (3.6±.38 versus 4.8±.38 seconds). This effect was not significant for percentage of errors ($_{ACC}F(1,18) = .33$, $p = .57$, $\eta_p^2 = .02$) There was also significant main effects of *Session* ($_{ACC}F(2,36) = 8.22$, $p = .001$, $\eta_p^2 = .31$; $_{RTs}F(2,36) = 10.49$, $p < .001$, $\eta_p^2 = .37$), and *Cycle* ($_{ACC}F(4,72) = 13.99$, $p < .001$, $\eta_p^2 = .44$; $_{RTs}F(4,72) = 13.41$, $p < .001$, $\eta_p^2 = .43$), with participants being both faster and less error prone with each session and with each cycle. Crucially, however, there was no significant interactions for *Condition* x *Session* interaction ($_{ACC}F(2,36) = .10$, $p = .91$, $\eta_p^2 = .01$; $_{RTs}F(2,36) = .32$, $p = .73$, $\eta_p^2 = .02$), *Condition* x *Cycle* ($_{ACC}F(4,72) = .32$, $p = .86$, $\eta_p^2 = .02$; $_{RTs}F(4,72) = .42$, $p = .80$, $\eta_p^2 = .02$), *Session* x *Cycle* ($_{ACC}F(8,144) = 1.06$, $p = .40$, $\eta_p^2 = .06$; $_{RTs}F(8,144) = .81$, $p = .59$, $\eta_p^2 = .04$), *Condition* x *Session* x *Cycle* ($_{ACC}F(2,36) = 1.33$, $p = .23$, $\eta_p^2 = .07$; $_{RTs}F(8,144) = 1.4$, $p = .18$, $\eta_p^2 = .07$).

EFFECT OF STIMULATION ON ACQUISITION



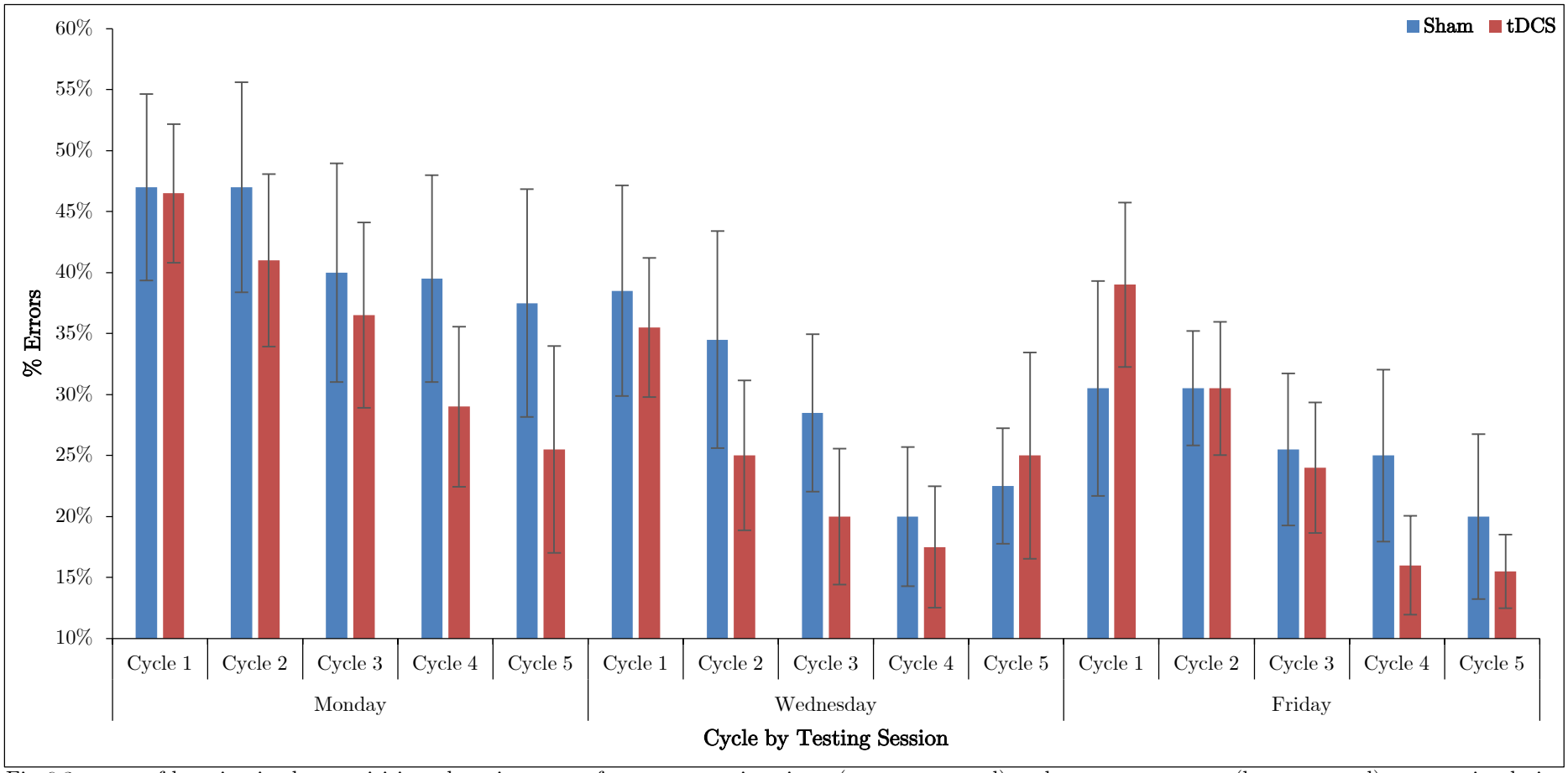
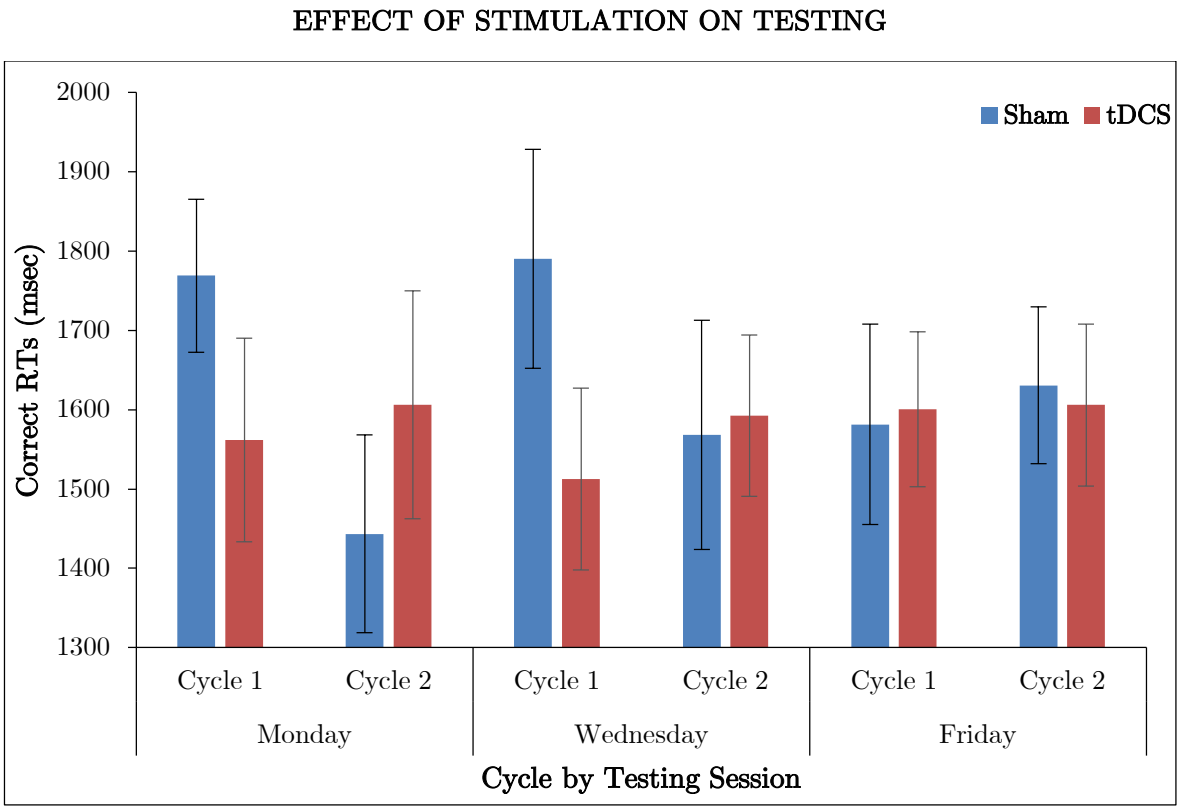


Fig 6.3 – rate of learning in the acquisition phase in terms of average reaction times (msec; top panel) and percentage errors (bottom panel) across stimulation conditions, cycles, and testing session. Error bars reflect standard error.

6.4.2.3 Testing Phase

I tested participants knowledge of items learned in the encoding and acquisition phase in terms of the percentage of errors and reaction times when responding to stimuli without feedback. Figure 6.3 shows that for reaction times were systematically faster during anodal tDCS compared to sham, and reaction times became gradually faster across sessions but only during anodal tDCS. I carried a mixed factors ANOVA, with *Condition* (Sham versus tDCS) as a between participants factor, *Session* (Monday, Wednesday or Friday), and *Cycle* (1 versus 2) as a within participants factor, and reaction times and percentage errors as the dependent measure. The results revealed no significant main effect of *Condition* ($_{ACC}F(1,18) = .03$, $p = .87$, $\eta_p^2 = .002$; $_{RTs}F(1,18) = 3.52$, $p = .08$, $\eta_p^2 = .16$), *Session* ($_{ACC}F(2,36) = 1.22$, $p = .31$, $\eta_p^2 = .06$; $_{RTs}F(2,36) = .78$, $p = .47$, $\eta_p^2 = .04$), and *Cycle* in terms of percentage errors ($F(1,18) = 1.87$, $p = .19$, $\eta p^2 = .09$), but there was a significant main effect of *Cycle* for reaction times ($F(1,18) = 1.42$, $p = .03$, $\eta_p^2 = .23$), with participant taking longer to go through stimuli in the first compared to the second cycle (1.10±.10 versus .98±.07 seconds). Crucially, there was no significant interactions for *Condition* x *Session* interaction ($_{ACC}F(2,36) = .34$, $p = .71$, $\eta_p^2 = .02$; $_{RTs}F(2,36) = .96$, $p = .39$, $\eta_p^2 = .05$), *Condition* x *Cycle* ($_{ACC}F(1,18) = .49$, $p = .49$, $\eta_p^2 = .03$; $_{RTs}F(1,18) = 1.42$, $p = .25$, $\eta_p^2 = .07$), *Session* x *Cycle* ($_{ACC}F(2,36) = 1.4$, $p = .26$, $\eta_p^2 = .07$; $_{RTs}F(2,36) = .54$, $p = .58$, $\eta_p^2 = .03$), *Condition* x *Session* x *Cycle* ($_{ACC}F(2,36) = .17$, $p = .84$, $\eta_p^2 = .01$; $_{RTs}F(2,36) = 1.00$, $p = .38$, $\eta_p^2 = .05$).



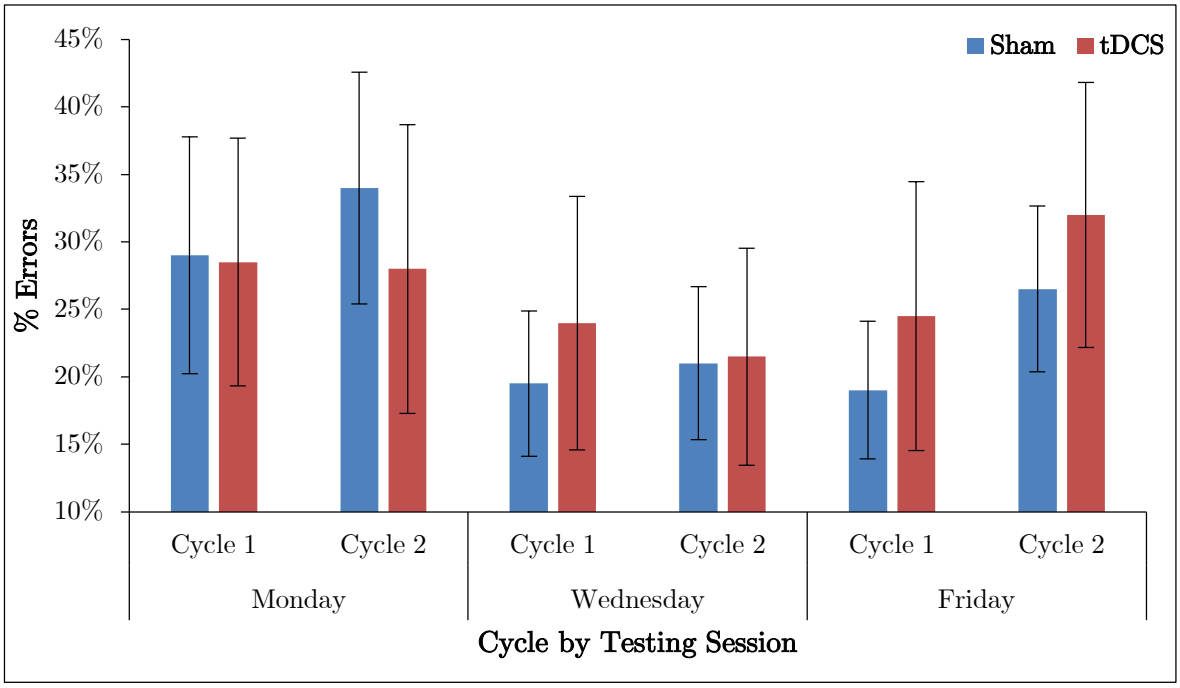


Fig 6.4 – performance in the testing phase in terms of reaction times (msecs) and percentage errors across stimulation conditions, cycles, and testing session. Error bars reflect standard error

6.4.2.3 Additional Consolidation Phase (~5mins).

I investigated the consolidation of items learned in previous sessions in subsequent sessions by asking participants to perform shortened versions of the acquisition and testing phases (referred to as re-acquisition and re-testing) before they learned new stimuli (see Figure 6.1 for more details). This analysis was only applicable to the Wednesday and Friday sessions. Figure 6.4 and 6.5 shows performance in re-acquisition and re-testing across stimulation conditions and cycles. For Friday's session, performance is broken down to see the consolidation of items presented on Monday and Wednesday. In all instances, there was no systematic effect of tDCS both in terms of the percentage of errors or reaction times. This was confirmed in a series of mixed factors ANOVAs, with *Condition* (Sham versus tDCS) as a between participants factor, *Cycle* (1 to 2 or 3, depending on the session and task) as a within-participants factor, and reaction times and percentage errors as the dependent measures.

Re-acquisition. For the ***Wednesday session,*** the results showed no significant main effect of *Condition* ($_{ACC}F(1,18) = .41$, $p = .53$, $\eta_p^2 = .02$; $_{RTs}F(1,18) = .21$, $p = .65$, $\eta_p^2 = .01$), but there was a significant main effect of *Cycle* ($_{ACC}F(2,36) = 7.42$, $p = .002$, $\eta_p^2 = .29$; $_{RTs}F(1,18) = 19.16$, $p < .001$, $\eta_p^2 = .52$), but no significant interaction for *Condition* x *Cycle* ($_{ACC}F(2,36) = 2.96$, $p = .07$, $\eta_p^2 = .14$; $_{RTs}F(2,36) = .04$, $p = .96$, $\eta_p^2 = .02$). For the ***Friday session***, I carried out analyses for stimuli learned on Monday and on Wednesday separately. In both analyses, the results showed there was not significant main effect of *Condition* (Monday stimuli: $_{ACC}F(1,18) = .00$, $p = 1.00$, $\eta_p^2 < .001$; $_{RTs}F(1,18) = .12$, $p = .74$, $\eta_p^2 = .01$; Wednesday stimuli: $_{ACC}F(1,18) = 2.09$, $p = .17$, $\eta_p^2 = .10$; $_{RTs}F(1,18) = .83$, $p = .34$, $\eta_p^2 = .04$), but there was a significant main effect of *Cycle* (Monday stimuli: $_{ACC}F(1,18) = 24.07$, $p < .001$, $\eta_p^2 = .57$; $_{RTs}F(1,18) = 28.82$, $p < .001$, $\eta_p^2 = .62$; Wednesday stimuli: $_{ACC}F(1,18) = 15.21$, $p = .001$, $\eta_p^2 = .46$; $_{RTs}F(1,18) = 16.82$, $p = .001$, $\eta_p^2 = .48$), but no significant interaction for *Condition* x *Cycle* (Monday stimuli: $_{ACC}F(1,18) = .18$, $p = .68$, $\eta_p^2 = .01$; $_{RTs}F(1,18) = .89$, $p = .36$, $\eta_p^2 = .05$; Friday stimuli: $_{ACC}F(1,18) = 2.79$, $p = .11$, $\eta_p^2 = .13$; $_{RTs}F(1,18) = .01$, $p = .95$, $\eta_p^2 < .001$)

Re-testing. For the *Wednesday session,* the results showed no significant main effect of *Condition* ($_{ACC}F(1,18) = .02$, $p = .89$, $\eta_p^2 = .001$; $_{RTs}F(1,18) = 1.26$, $p = .28$, $\eta_p^2 = .07$), but

there was a significant main effect of *Cycle* ($_{ACC}F(1,18) = 1.71$, $p = .21$, $\eta_p^2 = .09$; $_{RTs}F(1,18) = 5.41$, $p = .03$, $\eta_p^2 = .23$), but no significant interaction for *Condition* x *Cycle* ($_{ACC}F(1,18) = .35$, $p = .56$, $\eta_p^2 = .02$; $_{RTs}F(1,18) = .02$, $p = .88$, $\eta_p^2 = .001$). For the *Friday session*, I carried out analyses for stimuli learned on Monday and on Wednesday separately. In both analyses, the results showed there was not significant main effect of *Condition* (Monday stimuli: $_{ACC}F(1,18) = .04$, $p = .85$, $\eta_p^2 = .002$; $_{RTs}F(1,18) = .61$, $p = .44$, $\eta_p^2 = .03$; Wednesday stimuli: $_{ACC}F(1,18) = .06$, $p = .80$, $\eta_p^2 = .004$; $_{RTs}F(1,18) = .07$, $p = .80$, $\eta_p^2 = .004$).

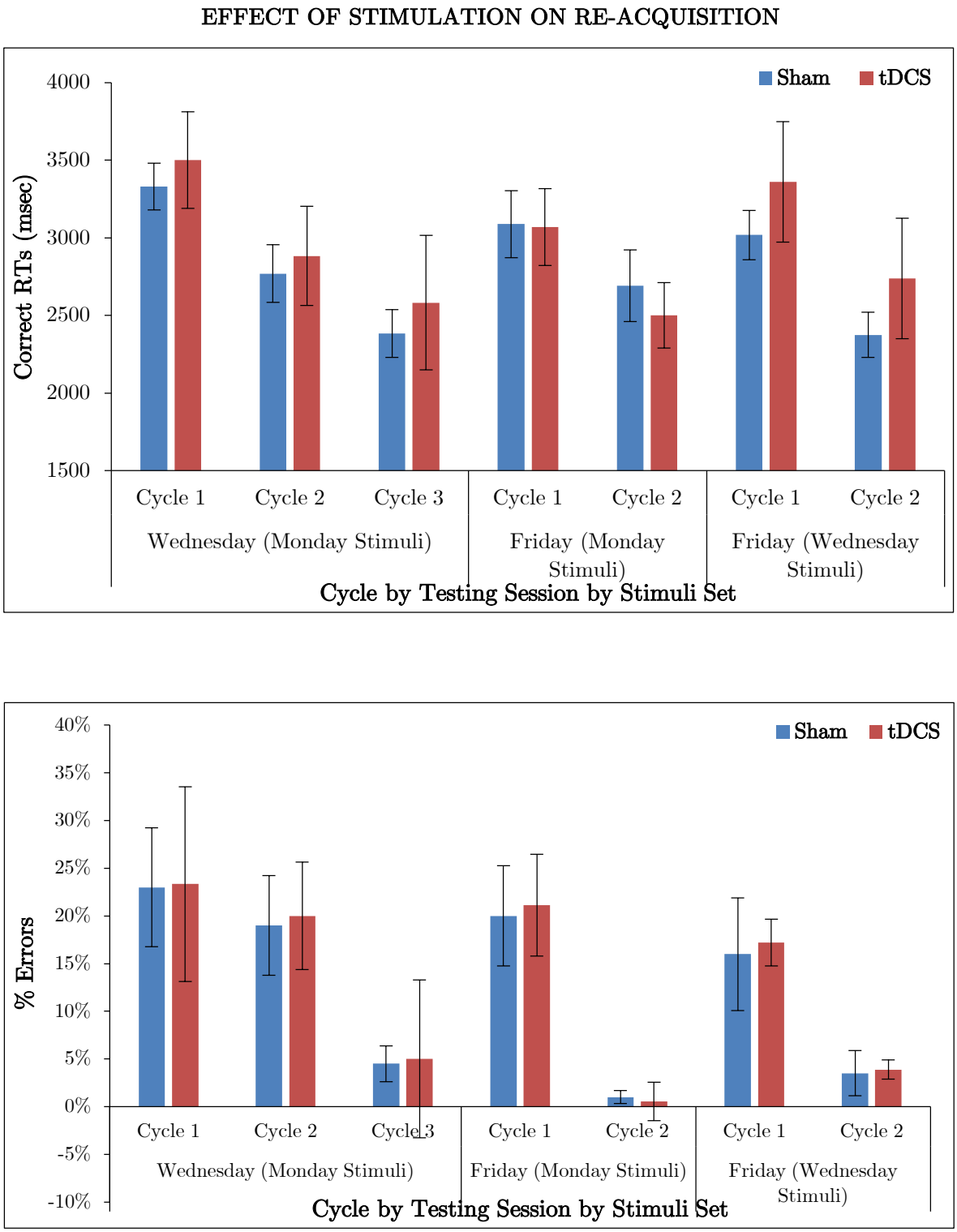


Fig 6.5 – performance in the re-acquisition phases in terms of correct reaction times (msec; top panel) and percentage errors (bottom panel) across stimulation conditions, cycles, testing session and stimuli set. Error bars reflect standard error

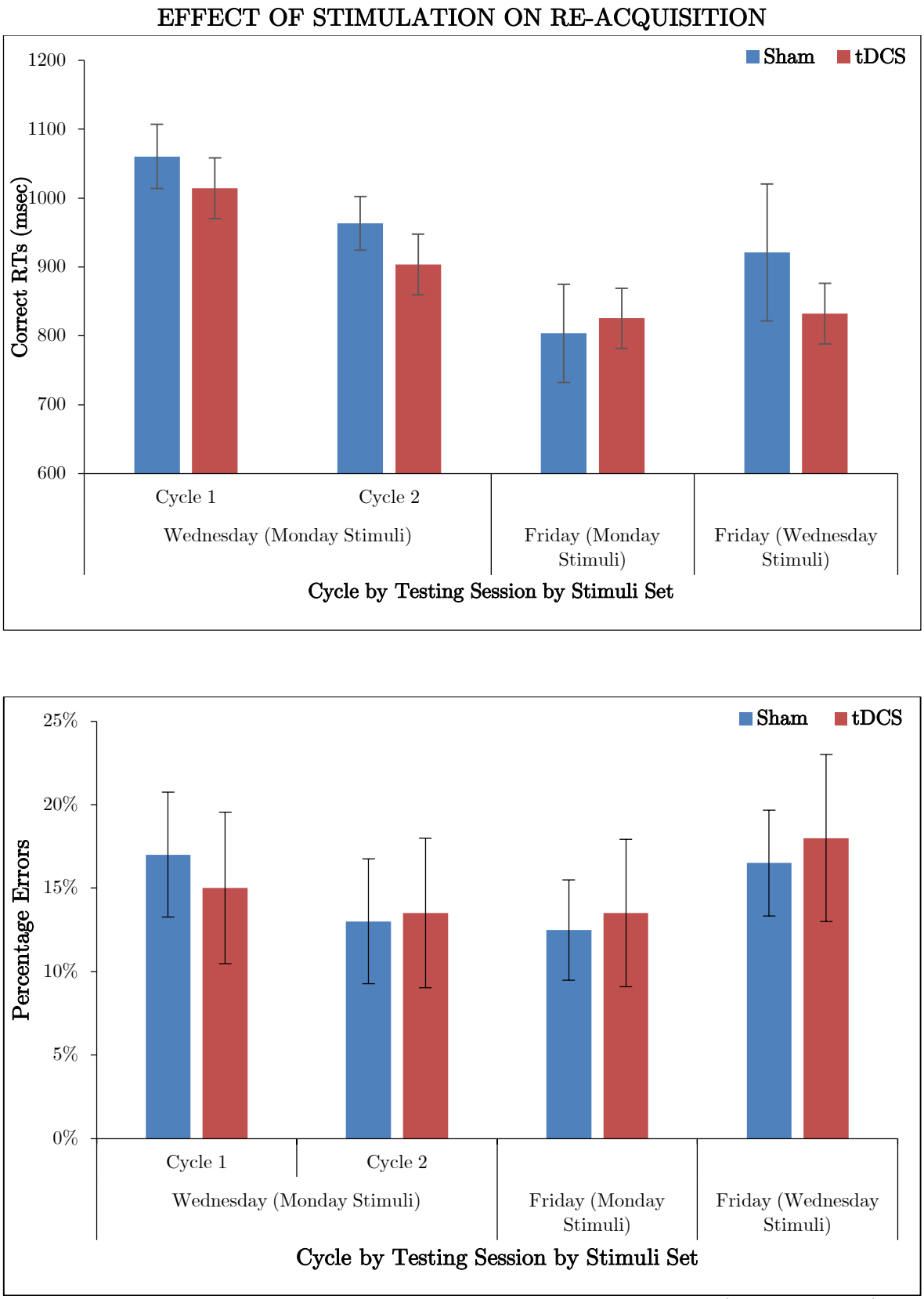
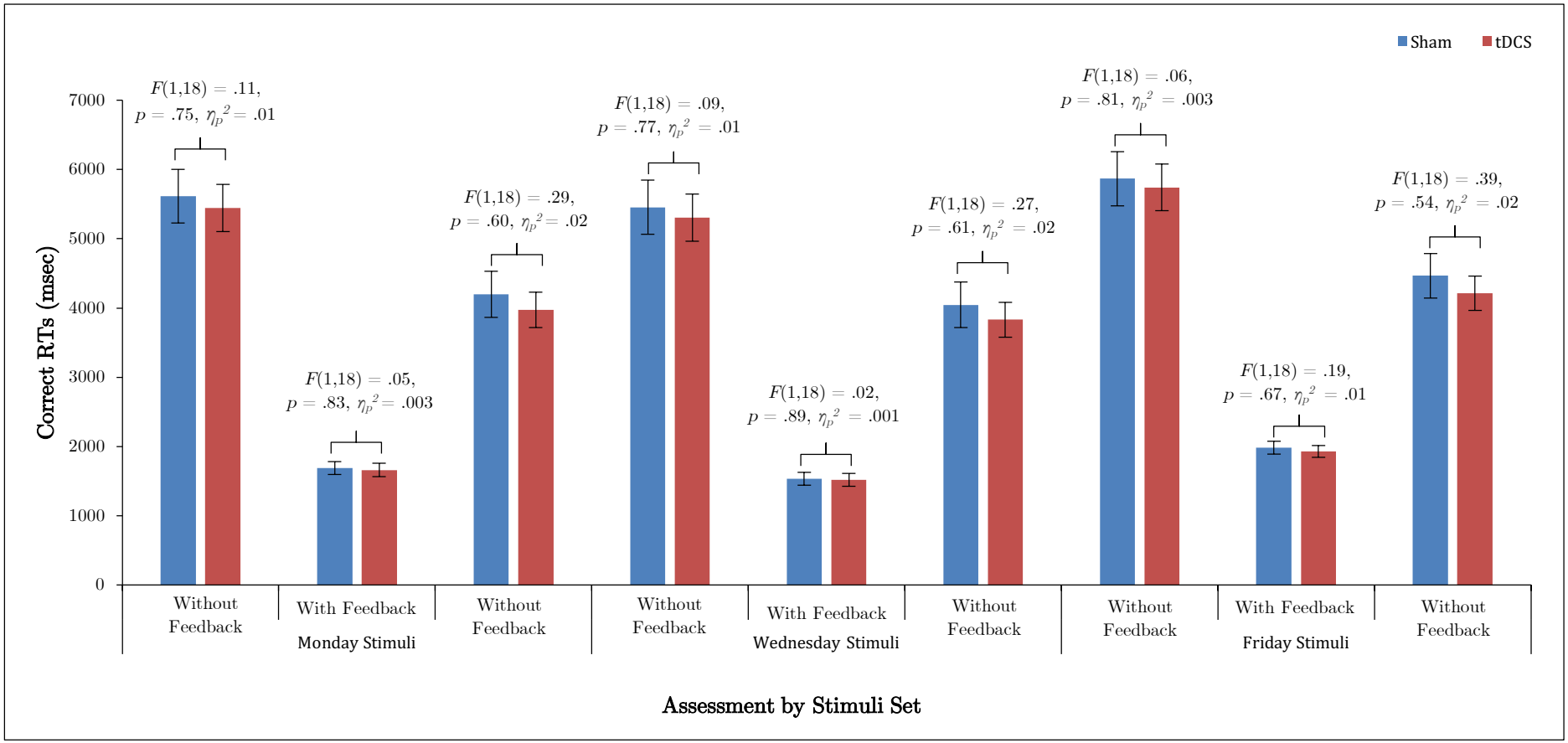


Fig 6.6 – performance in the re-test phases in terms of correct reaction times (msec; top panel) and percentage errors (bottom panel) across stimulation conditions, cycles, testing session and stimuli set. Error bars reflect standard error

6.4.2.4 Maintenance Session (~20mins).

I investigated the consolidation of items learned in previous sessions in a follow-up session carried out on the following Thursday, which involved assessing participants firstly with feedback, then without feedback and again with feedback. Figure 6.4.10 and Figure 6.4.11 show performance respectively in terms of reaction times and percentage errors across those three assessments, which performance broken down in terms of the stimuli set. Clearly, participants did not show any systematic effect of tDCS across any assessment (i.e., with or without feedback) and stimuli set. This null effect was confirmed in a series of one-way ANOVAs, with *Conditions* (Sham versus tDCS) as a between-participants factor. In one analyses, tDCS increased percentage errors, but given the number of null findings I do not put any weight on this significant effect.

FOLLOW-UP ASSESMENT



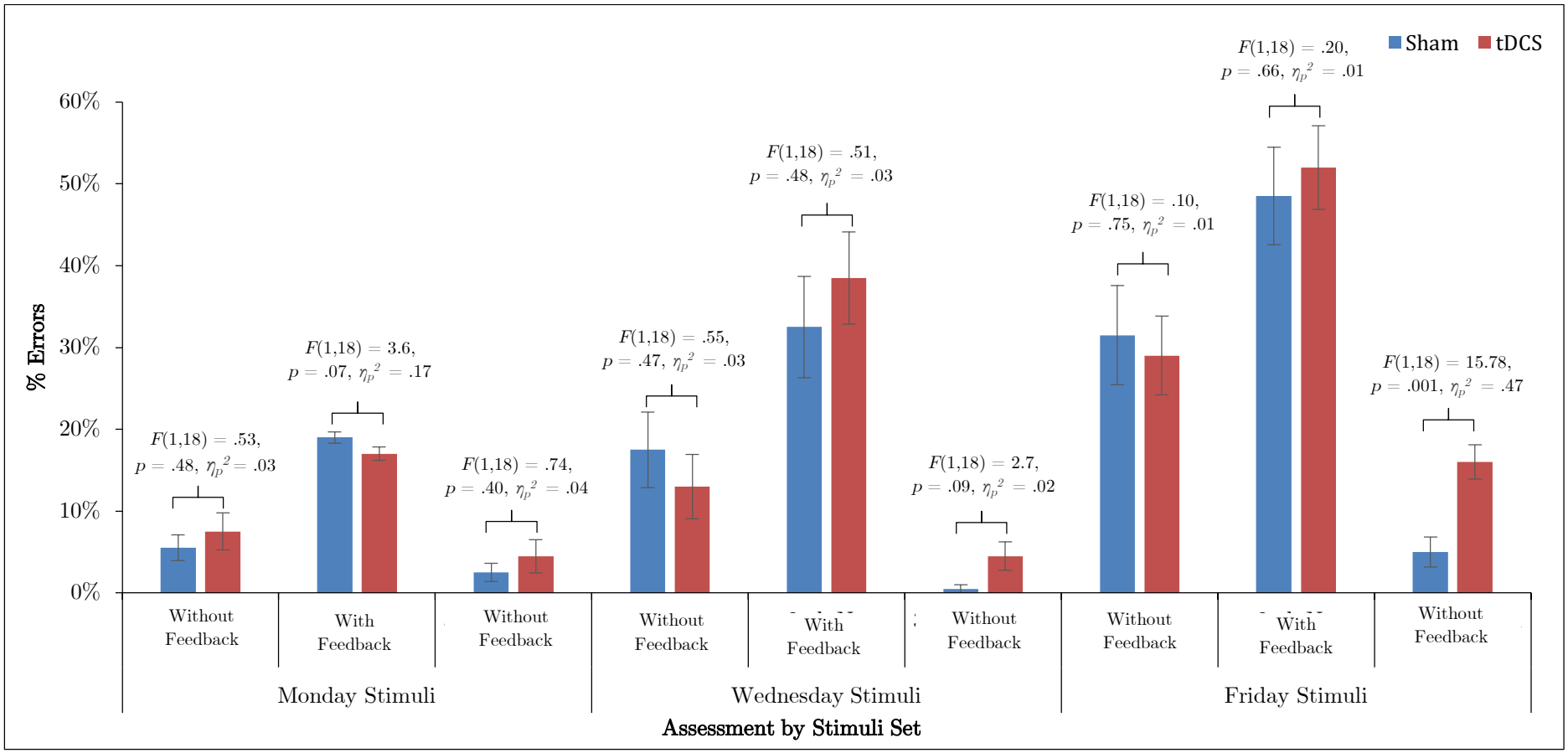


Fig 6.7 – performance in the follow-up session in terms of correct reaction times (msec; top panel) and percentage errors (bottom panel) across stimulation conditions, different assessments (first with feedback; second without feedback; third with feedback), and stimuli set. Error bars reflect standard error.

6.5 Discussion

When applied to healthy participants in a single session, tDCS is unlikely to lead to detectable changes in cognition, which may explain the null effects reported in my research findings and findings across the literature. Positive effects, however, may be more likely with repeated applications in conjunction with learning paradigms since the effects of tDCS can accumulate and can operate on neurons operating sub-optimally. In the present study, I investigated whether three alternate-day sessions (Monday, Wednesday, Friday) of anodal tDCS applied to the LIFG would enhance learning of a novel language. Consistent with this hypothesis, I found that in terms of the rate of learning across five cycles, participants were systematically faster and less error prone when given anodal tDCS compared to sham, but this effect did not accumulate significantly across sessions.

I note that these findings are preliminary and should be inferred with caution. Significant differences between conditions may simply be due to the between-participants design, which can allow a naturally occurring difference between groups to masquerade as an effect of tDCS – e.g., participants were simply faster in the tDCS condition regardless of any effect of tDCS. The small sample size, extreme variability in learning skills across participants, in addition to the variation introduced by the between-participants design, suggests that more participants will be needed to have enough power to detect positive effects of tDCS. At present I have collected data from twenty participants (ten in each stimulation condition), which is very short of the intended target of sixty participants (thirty in each stimulation condition).

What is clear from my preliminary results is that the paradigm I created works well. A chief concern was whether the task sufficiently challenged participants – to ensure that tDCS was exerting its effects on neurons operating at suboptimal levels. Examination of the percentage errors particularly during the learning phase shows that participants never reached ceiling even during the Friday session, where one would expect the effect of practice to be at its strongest. There was evidence of ceiling effects in testing, consolidation and maintenance, which also is to be expected given the repeated exposure to the stimuli. However, such variability can be easily reduced by adding more stimuli.

6.5.1 Conclusion

The effectiveness of tDCS may be determined by the integrity of neuronal activity. I propose that tDCS may operate preferentially in brain-damaged patients or during learning of a novel language (where there is suboptimal neuronal functioning) and/or during repeated application (where the effects of tDCS can accumulate). In the next section (6.6 Proposed future research), I set out an outline for future experiments, which will investigate these possibilities, thereby building on my work from previous Chapters, which attempted to verify the conditions in which single session tDCS may or may not be effective particularly in healthy participants. Preliminary findings from this proposed research demonstrate that significant effects of tDCS are possible in healthy participants during protocols that require participants to learn a novel language and undergo repeated sessions of tDCS stimulation. Although I do acknowledge that these effects may in actuality be driven by the lack of power and/or the between-participants design.

6.6 Proposed future research

I want to build on my research conducted in previous Chapters and in this Chapter to disentangle two possible, but not mutually exclusive, variables in explaining differences in the efficacy of anodal tDCS. These are: type of population and type of protocol.

6.6.1 Effect of population

In my first study, I aim to investigate effects of single-session tDCS in populations where cortical excitability may be suboptimal and to compare these with age matched controls and findings reported in previous chapters – i.e., populations where neuronal functioning is operating at optimal levels. I want to measure the effect of tDCS on picture naming and word reading in populations of healthy older participants and of brain-damaged patients, focusing specifically on aphasia sufferers given the reports of positive effects in this population. If tDCS effects are revealed when brain areas are working sub-optimally, I may find effects in these populations even if the paradigm does not involve learning and tDCS exposure is limited.

6.6.2 Effect of protocol

In my second study, I aim to compare the effects with healthy participants found in my previous experiments – i.e., effects in picture naming and word reading, which involve retrieval of consolidated representations – with effects in paradigms that probe learning of new words. If type of paradigm is crucial to demonstrate effects of tDCS, I may be able to demonstrate effects with learning and language stimuli, even in a healthy student population, in spite of having failed to see any effects previously on picture naming and word reading tasks. Moreover, I also intend to carry out this experiment with a multi-session design, with tDCS applied on alternate days of the week. If paradigm is inconsequential to securing an effect of tDCS, but the amount of exposure is crucial, then I will be able to demonstrate that effects in later stimulation sessions. If type of paradigm, however, is crucial as opposed to the number of stimulation sessions, then I would expect to see effects on the first day of stimulation. I am currently carrying out this experiment at the moment (see 5.4 Present research below for details and preliminary results).

6.6.3 Effects of protocol *and* population

Effects of population and protocol may combine to ensure positive effects of tDCS. In fact, the success of tDCS with aphasic participants may be due to all of the factors I have discussed: a) a clinical population where brain areas are operating sub-optimally; b) learning paradigm; and c) repeated tDCS applications. In my third study, I want to assess combined effects of population and paradigm by considering effects of tDCS on learning in populations of older adults. Positive effects in the second or third study will open the way to assess differences in other populations where brain areas may operate sub-optimally (e.g., assessing tDCS effects in populations which are stressed or fatigued).

Chapter 7: Discussion & Conclusion

7.1 Thesis Summary

In recent years, there have been calls for the re-evaluation of tDCS despite several thousand papers reporting positive effects across a broad range of cognitive functions. This sea change has been brought about because of evidence that effects of tDCS are largely exaggerated; are from studies using samples with low evidential value; and are difficult to reproduce, especially in studies using conventional stimulation parameters – i.e., single applications of anodal tDCS with healthy individuals. This thesis set out to contribute to this question by investigating the efficacy of conventional stimulation protocols with reference to popular explanations for the poor reliability of tDCS in healthy participants (i.e., differences in stimulation parameters; cortical excitability; cognitive load; and statistical power), and with a specific focus on modulating performance on word production, working memory, and language learning.

First, in Chapter 2, I carried out four experiments in which I applied anodal tDCS using conventional stimulation parameters for targeting key language areas – i.e., left inferior frontal gyrus (LIFG) and left posterior temporal lobe (LpTG) – during picture naming and word reading tasks. Unfortunately, our results failed to show an overall effect of tDCS. Null effects were also found in further analyses considering semantic interference (a proxy for changes in cortical excitation), individual variation in response to tDCS effects, and when analysing naming speeds for difficult items. These null results were subject to a commentary piece, in which limitations to my investigation were highlighted. However, in my rebuttal, I demonstrated that these limitations had no merit.

Second, in Chapter 3, the surprising null effects reported in Chapter 2 prompted an investigation into the foundational claim that tDCS can modulate word production. This led to a systematic review and meta-analysis pooling data from studies measuring single sessions of tDCS on healthy participants using picture naming and word production tasks. Across several meta-analyses, there was no evidence that tDCS significant modified performance. Furthermore, despite these null effects, publication bias was detected that

indicated a bias towards reporting positive effects in studies using conventional stimulation parameters – i.e., conditions in which bias would be expected.

Third, in Chapters 4, I moved focus to assess whether anodal tDCS applied to the LIFG could modulate performance on verbal fluency and working memory tasks, areas where there are more positive effects reported in comparison to word production research. Unfortunately, again, I failed to show any significant effect of tDCS on overall performance. Additional analyses considered conditions of increased cortical excitation due to demands on executive selection abilities (i.e., switching versus clustering in verbal fluency; interference effects in working memory tasks), participant variation in response to tDCS effects, and with difficult items. All of which showed no effect of tDCS.

Fourth, in Chapter 5, I carried out a replication of a previous study showing large positive effects across verbal fluency task (i.e., Cattaneo et al., 2011) in addition to a systematic review and meta-analysis to quantify the effect of tDCS on verbal fluency tasks in studies using healthy participants. In the replication I failed to find a significant effect, despite following closely the method outlined by the original study. In the meta-analysis, I found a significant positive effect of tDCS, with semantic fluency tasks being particularly sensitive to the effects of tDCS. However, on close inspection, I found publication bias, and that the significant effect was driven by a minority of underpowered studies reporting exceptionally large effects, one of which my replication failed to replicate.

Finally, in Chapter 6, I outlined a plan to investigate potential conditions in which tDCS may be effective. These include the modulation of unstable representations, such as in the case of novel language acquisition, with repeated application of anodal tDCS in both healthy and aphasic patients. I also report preliminary work that demonstrates that tDCS may be effective in these conditions, especially on the third session of tDCS – however this is only pilot data using healthy participants.

Thus, in summary, the extensive findings reported in this thesis show that after many and diverse attempts (all geared toward increasing the chances of finding a significant outcome), I found no evidence in favour of tDCS. These null findings specifically apply to conditions

in which anodal tDCS was applied to the LIFG and LpTG in single sessions whilst healthy participants performed word production and working memory tasks. The preliminary work conducted on multi-sessions on novel language acquisition does show some positive evidence of an effect of tDCS, but this work is still on going. This thesis, therefore, makes clear the stark disparity between the popularity of tDCS and the lack of evidence when systematic investigations are carried out, thereby raising important concerns about the quality and future direction of tDCS research and possible solutions to errors in what is considered normal practice.

7.2 Future Directions: Questionable Practices & Solutions

Following my extensive reading of the literature and work reported in this thesis, it is clear to me that tDCS research needs far more rigor. Below I go through questionable research practices that others have reported and I have seen first-hand, along with potential solutions to resolve these issues.

7.2.1 Questionable Research Practices

In a series of studies that surveyed researchers in the field, a stark disconnect was revealed between the widely reported positive effects and the first-hand experience of researchers using tDCS. In one study, only 19 to 33% of respondents considered tDCS to be "mostly effective" (19-33%; 28-42%, "partly effective", 2-5%, "ineffective"; 2-13%, "absolutely effective"), whilst another found that only 50% of researchers could successfully reproduce findings from previous studies. Worst still were the comments from respondents. Questions were raised about tDCS efficacy for cognitive enhancement in healthy participants, along with doubts about the credibility of the field (e.g., one respondent stated that, 'I think there is a huge publication bias in this field and, in my opinion, the positive results of tDCS are highly overestimated'; see Riggall et al., 2015). Respondents knew of researchers who increased the chances of a significant outcome by adjusting statistical analysis and/or deliberately reporting a selection of all study outcomes (see Héroux et al., 2017; Riggal et al., 2015), and that researchers who considered tDCS to be central to their research (compared to those who did not) rated its efficacy to be higher, and a similar pattern was found for junior researchers (see also, Walsh, 2013).

On top of this, data mining appears to be epidemic in tDCS research. Medina and Cason (2017) carried out a p-curve analysis, which analysed the frequency of reported significant *p*-values between 0 and .05 (i.e., the threshold of significance) across tDCS studies measuring cognitive effects. The assumption was that – if there is an effect of tDCS – the distribution should be positively skewed, with a majority of reported *p*-values leaning towards, for example, .00 or .01 rather than .04 and .05. In the presence of data mining, however, one should expect to find a greater negative skew, where researchers have stopped collecting participants or exploratory analyses once they have passed the critical $p = .05$ threshold (i.e., data mining or *p-hacking*). The latter is in fact what Medina and Cason discovered across a pool of 30 studies they collected following a systematic search, and 22 studies from a previous meta-analysis (i.e., Mancuso et al., 2016). What is more troubling, however, of those studies which reported p-values close to 0, many had not been replicated, such as Cattaneo et al. (2011), which I failed to replicate in Chapter 5. Similarly, in my other work I found evidence of similarly questionable research, namely in terms of effects arising from underpowered studies and publication bias.

Low power raises the likelihood of a null effect, but in the case of a significant effect the effect is likely to be a false-positive and is inflated because the study is only powered enough to detect large effects (Button et al., 2013; Button, Bal, Clark & Shipley, 2016; Forstmeier, Wagenmakers & Parker, 2016; Ionnides et al., 2005). The problem with low power is inherent in tDCS research. The recruitment of 20 participants (or often fewer) is generally regarded to be an acceptable sample size, with the typical power achieved across tDCS studies being 4 to 14% (see Medina & Cason, 2017). In Chapter 3 and 5, I present evidence of exceptionally large effect sizes reported in word production tasks, which were from small underpowered studies. More importantly, however, in Chapter 5, I show that these effects can drive a significant summary effect size in meta-analyses. Replications are one means to verify findings, and – as was shown in Chapter 5 – my attempt to replicate failed, and a growing body of evidence shows that other attempts invariably fail or at least show a reduced effect compared to the original (e.g., Brückner & Kammer, 2016; Emmerling et al., 2017; Horvath, Vogrin, Carter, Cook & Forte, 2016; Jalali, Miall, & Galea, 2017; Nilsson, Lebedev, Rydström & Lövdén, 2017; Spielmann et al., 2017; Vannorsdall et al., 2016). Thus, though

I acknowledge that the effect of tDCS may be large enough to be detected by samples as small as twenty participants, there is little evidence to justify this.

Another issue is publication bias – i.e., the reluctance of journals and researchers to publish null results, favouring instead novel or positive findings (Martin & Clarke, 2017). I found evidence of publication bias in both Chapters 3 and 5, and my attempt to produce a conceptual and direct replication in Chapters 2 and 5 are further evidence that null effects are likely going unreported. Bias in approaches to publication will inevitably hold a field hostage to a biased selection of data of all studies carried out, which in the case of tDCS – where there is potential bias toward reporting positive effects – may present its effects as being unjustifiably reliable. What is worse, however, the estimates of bias in tDCS potentially down play the problem. In new and rapidly expanding research areas, where funding is competitive and the appetite to confirm previous findings is lacking, as is with tDCS research, the likelihood is that publication bias is widespread.

Finally, career ambition is a plausible account for why these questionable research practises occur, which would be consistent with explanations given elsewhere in other areas of science where similar practices have been reported (Romain, 2015). However, unlike most other research areas, the negative repercussions for tDCS could be greater if the quality of science is compromised to inflate the significance of results. The problem as been discussed candidly by Vincent Walsh (2013), an expert in non-invasive brain stimulation: 'when my friends and colleagues say that "tDCS is a non-invasive brain stimulation (NIBS) neuromodulatory technique, whose clinical applications to treat pathological neuropsychiatric conditions are rapidly growing [Santarnecchi, Feurra, Galli, Rossi, & Rossi, 2013]." I think they fall into a language trap (in which we all find ourselves) of confusing claims with reality. ... I am all for hope, but when it crosses the line into faith, it becomes an unthinking vehicle. ... [One] consequence of the hype is that the noise may mask important findings. We saw the effects of this with depression and TMS, the advance of which was slowed by premature claims and masked by claims about the utility of TMS in just about every neurological and psychiatric condition....'. What is worse, this problem is compounded by the fact that the field of tDCS uniquely incentivises researches to inflate the significance of tDCS effects in the service of securing one of the many lucrative roles on advisory boards for various stimulation device

manufacturers or to help get the edge over competition for funding (particularly funding from charities who are not experts in tDCS and would therefore be easily led by research suggesting that tDCS might be an alternative treatment for treatment-resistant disorders). This is also pointed out by Walsh (2013), who goes on to say: "it is hard to find a colleague working on clinical or enhancement aims who does not have an industry consultancy, shares in a company or a patent filed to protect their interests (50% of my own salary is funded by the Royal Society to work with industry – The Magstim Company, for example)'. It is hard not to see how this can have potential costs later down the line, not only to the confidence in the rigor of tDCS research, but also the wasted time and effort pursuing effects of tDCS that turn out to be not as solid as previously thought, and the knowledge lost by not checking or being transparent about the robustness of one's findings.

7.3.2 Solutions

Recent years have seen initiatives to introduce rigor into various scientific fields, with the Open Science Framework being a leading light (Munafò, et al., 2017). These initiatives have the central aim of making research findings transparent, which is now an achievable goal thanks to advances in the storage and sharing capabilities of modern computers. The ambition is to increase the reliability of scientific research, and the quality of tDCS research will be much improved if it were to adopt this ambition. Below I recommend possible tools one can use to improve the quality of tDCS research.

7.3.2.3 Greater Transparency

There should be more pressure for greater transparency so the reader is made aware of the vested interests of the researcher(s) and those who fund them. Declaring conflicts of interest is considered best practice. It is mandatory in many university institutions and journals, and carries with it severe punishments for not declaring conflicts. This can therefore be an effective tool with wide utility, although there is considerable variation in adopted policies, and not all researchers declare everything, or they find ingenious ways to circumvent the rules (see Romain, 2015). More importantly, however, it does not fully address one area where the impact of protecting one's own interests could be felt, and where the impact may be most negative, namely peer-review. Peer-review is single-blinded in most journals, whereby the reviewer knows the names of the authors, thereby leaving open the possibility

for reviewers to practice cronyism, or a favourable (or unfavourable) decision to publish based on the author's previous work rather than the scientific merit of the submitted manuscript under review, as seen elsewhere in science (Martin & Clarke, 2017). The situation is compounded when you consider that many opinion formers in tDCS are editors on journals in which tDCS studies are predominantly published. Editors yield the power to reject a submitted manuscript even before it is selected to go to peer review, and can therefore exercise this power in cases where a manuscript or its author throws shade on work on which some of the editor's current and future career success is based. This naturally makes the publication of null findings even harder than it already is. One solution might be to anonymise the names of the authors on the submitted manuscripts (i.e., double-blinded peer review). However, tDCS research is a relative small research field, where research is typically aired at a handful of specialised conferences, and a small number of labs work in particular research domains, meaning that identifying the authors would not be hard to achieve. One alternative solution would be to have greater transparency. Journals like *Frontiers*, for example, permits readers to see the editor and the reviewers of each published manuscript, which would no doubt help the reader weigh the scientific worth of a paper by how rigorously it was reviewed. This open approach would be particularly effective in the case of Registered Reports and 'Results Free' Peer Review, which I go on to next.

7.3.2.1 Pre-Registration, Registered Reports, 'Results Free' Peer Review

The most popular recommended tools against publication bias and data-mining are pre-registration (PR) and registered reports (RRs; for more discussion, see Munafò, et al., 2017). PR involves registering a summary of a study protocol including planned statistical analysis before data collection to an online platform (e.g., Open Science Framework, retrieved from https://osf.io/; AsPredicted, retrieved from https://aspredicted.org/), with a commitment to publicising results. Thus, the public a) have access to a complete record of all current research (reducing publication bias), and b) can distinguish between pre-specified and exploratory analysis (reducing data mining). RRs use a two-stage peer review process. First, before data collection, a detailed protocol is submitted to a journal that accepts RRs[12],

12Over 70 journals accept RRs, some of which are high impact journals, such as Cortex (retrieved from https://cos.io/rr/?_ga=2.41992997.1441516866.1506449919-496589304.1502031854#journals).

reviewers then peer-review the protocol before agreeing to publish, providing the protocol has scientific merit and is strictly followed. Second, the submitted manuscript is peer-reviewed and published, if the protocol was followed and inferences about results are justified. Thus, reviewers cannot be biased by results, data mining is constrained, and researchers can carry out research in the safe knowledge that null findings can be published (Munafò, et al., 2017).

However, PR and RRs come with problems that one must bear in mind with reference to tDCS research. In regards to PR, authors can choose whether to publicize their pre-registered study, and PR does not tackle journal bias (i.e., the tendency to favour the publication of positive results), which has a bearing on whether researchers prioritise the publication of null results. This is addressed in RRs – since the decision to publish is taken before the results are known. However, RRs take several months to complete before data collection even begins, and the rejection rate before the protocol reaches the first round of peer-review is not well documented. Many claim this will incentivise the submission of high quality work to avoid delays and/or rejection, but I fear it may encourage researchers to choose the traditional 'post study' peer review publication route, especially given the lack of uptake of RRs across the field in 'go to' journals such as *Brain Stimulation* and since tDCS experiments take just three to four weeks from data collection to write up (see Mancuso et al., 2016; Minarik et al., 2016). A potential solution, however, is a gradual transition towards RRs. PRs may be one option, but the 'results free' format may be another, which requires journal reviewers to review a manuscript submitted without results or discussion (which are seen at a later stage to check if methods were followed and/or if minor revisions are needed). Thus, unlike RRs, blinded peer review offers a potentially faster and familiar route to most researchers. Yet, like RRs, by not disclosing the results, reviewers cannot be biased nor are researchers incentivised to publish positive effects. Moreover, however, because publication hinges on the quality of the seen parts of the submitted manuscript (e.g., introduction, rationale, method), and close adherence to the method in the results section, the quality and rigour of submitted work should be high. The one downside to results free peer review, however, is that it has been piloted in only a limited number of journals (see Button et al., 2016).

7.3.2.2 Replication

Reproducibility is a cornerstone of science, and the reproducibility of an effect can be tested with either direct or conceptual replications. *Direct replications* try to replicate previous work by following strictly a methodological procedure used by a previous study. *Conceptual replications* are novel studies that tests a hypothesis or result reported in previous work (Harris, 2017; Schmidt, 2009; Simons, 2014). If nothing else, the priority of a PhD student is to repeat their own and other researchers work. The results from these replications may prove invaluable in the early stages of a PhD, where the plan for subsequent investigation is mostly formulated. I can say from personal experience that the failed direct and conceptual replications of previous studies measuring effects on word production led me to return to first principles, which took the form several meta-analyses and one replication which now question the efficacy of tDCS, at least in the conditions I explored.

7.3.2.3 Collaboration

Replication can be costly both in time and money. In tDCS research this is particularly true since my work and others make clear that replications need to be large scale to combat the problem with low statistical power, placing more strain one usually scant resources. Collaborative research projects can be effective when trying to replicate results, because they allow different research groups to pool resources and boost sample sizes. Collaboration has been effective in the 'Many labs replication project' (or MLRP; https://osf.io/wx7ck/, see Klein et al. 2014), in which 36 labs replicated 13 widely reported effects in psychology research, and found that 10 were replicable.

7.3.3 Expanding frontiers

The results reported in this thesis can speak to only those conditions in which I investigated. These include single session anodal tDCS applied to the LIFG and LpTG during word production and working memory tasks, with preliminary research conducted on repeated applications over alternate days. These conditions were chosen because they speak to a large portion of the tDCS literature, but there is still much to be learned about the effects of tDCS in other conditions and non-invasive brain stimulation techniques. I therefore recommend future avenues of research that should receive attention.

7.3.3.1 Sample Population.

As noted in Chapter 6, the outcome of tDCS may depend on the population, with healthy participants being potentially less amendable given that cognitive performance is already at optimal levels. Therefore, this thesis reserves judgment regarding the effectiveness of tDCS on younger, older, and patient samples and cognitive domains. Moreover, more efforts should be taken to expand work to investigate its efficacy in patient populations, where the application of tDCS will have the greater impact on people's lives. Currently, the majority of work carried out is conducted on healthy participants. This is perhaps understandable given that patients are rare, especially stroke patients, yet this problem might be stymied if researchers are encouraged to collaborate.

7.3.3.2 Selected Protocol

To constrain variation in outcomes across studies it is important that future research carries out a systematic exploration within the conventional parameters described in Chapters 1. I admit that the focus of this thesis was limited because I focused only on anodal tDCS, and targeted the left frontal regions. As has been mentioned through, this decision was based primarily on the fact that a majority of studies have targeted frontal regions with anodal stimulation, and within the scant evidence with cathodal tDCS or tDCS of either polarity of different brain regions one mainly finds inconsistent results. This of course does not mean that tDCS may not be effective in these conditions, and that we should not consider them to be possible avenues of future research. It should be taken as a challenge for future research.

7.3.3.3 Widening the tool kit

Most tDCS devices are capable of administering other forms of non-invasive brain stimulation that apply an alternating as opposed to a direct current. Research using these techniques is growing, and although evidence is still coming in, studies published thus far show encouraging results, and therefore should be explored as an alternative to tDCS given the shortcomings discussed in this thesis. Additionally, attempts are being made to refine current tDCS protocols, mainly with the use of computational models of current flow which may inform the planning of stimulation protocols and the development of stimulation montages, thereby optimising the neurophysiological effect of tDCS and therefore improve

chances of influencing cognition. Although tDCS was the focus of this thesis, it is worth exploring these new research areas in more detail.

Transcranial alternative current stimulation (or tACS). In tDCS, a constant current flows from one electrode (the anode) to the other (the cathode). By contrast, tACS administers pulses of current the direction of which alternates regularly between the two electrodes – thereby changing which electrode is functionally the anode and the cathode at regular intervals. The rate at which the current alternates usually follows a biphasic sine wave, which, depending on the frequency, can modulate cortical excitability. For example, following 10 versus 20 Hz (alpha versus beta frequencies, respectively) of AC stimulation over the primary motor cortex, cortical excitability respectively decreased (Antal et al., 2008) and increased (Feurra et al., 2011) as measured by TMS evoked MEP amplitudes, with affects lasting 30 to 70 mins post stimulation cessation (Zaehle, Rach & Herrmann, 2010; Kasten, Dowsett & Herrmann, 2016). These after effects suggest that the tACS and tDCS probably share the same basic mechanism of action, namely the modulation of resting membrane potentials that then lead to plasticity changes which prolong excitability changes post stimulation. Similarly, like tDCS, tACS effects may be intensity dependent, with lower intensities leading to inconsistent outcomes. For example, 0.4mA lead to inhibition of motor cortical excitability, with no effect on 0.6 and 0.8mA, yet 140 kHz or 1-5kHz tACS at an intensity of 1mA induces large increases in MEP amplitudes (Moliadze et al., 2012).

The main appeal of tACS over tDCS is that when the frequency of pulses are within conventional EEG frequency ranges (i.e., 0.1–80 Hz), it has the potential to externally modulate ongoing oscillatory activity (for review: Herrmann, Rach, Neuling, & Strüber, 2013). It is widely thought that higher-order cognitive functions rely on cortical oscillatory activity (Wang, 2010; Donner and Siegel, 2011; Schutter & Wischnewski, 2016), which are instrumental for communication between brain areas (Thut & Miniussi, 2009) and the integration of bottom-up and top-down information (Engel, Fries, & Singer, 2001; Varela, Lachaux, Rodriguez, & Martinerie, 2001). Thus, by externally modulating oscillatory activity, tACS can help our understanding of the relationship between oscillatory activity and cognition, in addition to providing a biologically plausible means to modulate higher cognitive functions that operate at different frequency bands and to improve various

neurological disorders characterised by dysfunctional oscillatory activity, such as Epilepsy, ADHD, Parkinson's, Alzheimer's and Schizophrenia or (Herrmann and Demiralp, 2005; Uhlhaas and Singer, 2006, 2012). Although relatively few studies have explored the effects of tACS compared to tDCS, of those studies carried out positive effects have been reported, albeit with some conflicting results.

In one study, tACS at 5hZ (theta) over the left parietal region or left frontal region led to gains on difficult items in two fluid intelligence tasks, which the authors ascribed to gains in working memory storage. These gains we also associated with an increase in frontal theta and a decrease in posterior alpha power (Pahor & Jaušovec, 2014). However, in another study, gamma-tACS over the left middle frontal gyrus reduced speed of responses in a similar measure of fluid intelligence. However, alpha-, theta-, and beta-tACS had no effects compared to sham stimulation. The same authors found benefits of gamma-tACS on logical problem-solving, especially in participants with slower baseline response times, suggesting that like tDCS tACS effects might be determined by baseline ability (Santarnecchi, et al., 2013, 2016). Similarly, in a Go/NoGo task, where participants must respond or withhold their response depending on a cue presented on a computer screen, it was revealed that whilst 70 Hz tACS improved accuracy during GO trials there was no related effect during NoGo trials where there is a greater need to exert inhibitory control (Joundi, Jenkinson, Brittain, Aziz & Brown, 2012). In a visual memory-matching task, responses were faster with synchronised theta-tACS and slower and less accurate with desynchronised theta-tACS (Polanía, Nitsche, Korman, Batsikadze & Paulus, 2012), whilst risk-taking behaviour in a decision-making was increased with theta-tACS when applied to the left dlPFC (Sela, Kilim & Lavidor, 2012), and, finally, 40 Hz gamma-tACS facilitated disengagement and reorientation in an attentional task (Hopfinger, Parsons & Fröhlich, 2017). Findings with tACS are, therefore, promising. However, perhaps the most attractive aspect of tACS is that unlike tDCS the dose can titrated to target the oscillatory activity of individual participants. One study, for instance, used EEG to record individual alpha oscillatory activity to inform the frequency range of the applied tACS to significantly boost individual alpha power (see Zaehle et al., 2010), with effects lasting up to 30 minutes post stimulation cessation. Thus, by being able to modify parameters to a specific individual, it reduces variation in response to tAC stimulation, a key limitation with tDCS.

Transcranial random noise stimulation (or tRNS). tRNS was developed to externally introduce noise to desynchronize normal and pathological rhythms of neuronal networks by administering pulses of alternating current at different amplitudes and frequencies (Terney, Chaieb, Moliadze, Antal, & Paulus, 2008). This contrasts to tACS where typically a single frequency is applied at a constant intensity. Very few studies have explored tRNS, yet of those studies that have been carried out promising findings have been reported. For example, 10mins of tRNS can significantly increase motor cortex excitability in healthy participants at high frequencies (i.e., 100-640 Hz; Terney et al., 2008), with after effects lasting for roughly 60 minutes, suggesting neuroplasticity changes similar to tDCS and tACS. Similarly, studies using tRNS and fMRI indicate a reduction in BOLD following stimulation, suggesting a reduction in energy consumption, an indication of increased *efficiency* of targeted neuronal populations (Chaieb et al., 2009). tRNS is also perceived to be more comfortable than tDCS (Moliadze, Antal & Paulus, 2010), which has implications for how well the experimenter and participant are blinded – thereby overcoming growing concerns about problems with blinding in tDCS research (Ambrus, Paulus, & Antal, 2010; Horvath, 2015). Evidence of cognitive effects are limited and contradictory, however. When applied over the dorsolateral prefrontal cortex (dlPFC), null effects were seen an *n-back* task (Mulquiney et al., 2011), errors increased on a probabilistic classification task (Ambrus et al., 2011), whilst learning improved on a mental arithmetic task with effects persisting up to 6 months after stimulation (Snowball et al., 2013). Finally, in another study, perceptual learning was improved when tRNS was applied over the visual cortex (Fertonani, Pirulli, & Miniussi, 2011). Thus, it is unclear how effect tRNS is when modulating cognition, but future research should address this gap in our knowledge.

A research area that also needs attention is the optimal parameters in which to apply tRNS, which to date is under investigated. Some evidence suggests that, like to tDCS, tRNS applied for 5 or 6 minutes (at frequencies of 100–640 Hz) leads to a significant increase in motor cortical excitability, but no effect was seen with 4 minutes of stimulation (Chaieb, Paulus, & Antal, 2011). More importantly, however, research needs to elucidate the physiological mechanisms of tRNS. Like tDCS, tRNS is thought to modulate the activity of sodium ion channels, thus leading to the induction of plasticity changes (Chaieb et al. , 2009; Terney

et al., 2008; Paulus, 2011) or like tACS it might increase how sensitive neuronal networks are to being modulated (Francis et al., 2003). It has, however, been demonstrated that tRNS effects are generally excitatory particularly with faster oscillating frequencies, where depolarizition of neurons would occur regardless of current flow direction (see Fertonani et al., 2011; Terney et al., 2008). Moreover, because tRNS disrupts neuronal networks at random, it is less likely to engage homeostatic mechanisms, which typically respond to prolonged, tonic disruption to the neural networks – a scenario more likely under tDCS. The implication is that, in comparison to tDCS, tRNS may evoke larger excitatory effects, leading to greater chances of detecting changes in cognitive performance. However, this assumption has not been thoroughly tested.

Current modelling. In a typical two-electrode arrangement, direct current flows from the anode to the cathode. The path of current flow is thought to be diffuse and therefore complex, with its path largely determined by the conductive tissue of the skull and brain. For example, cerebral spinal fluid (CSF) is a highly conductive fluid found throughout the brain, diverting (or "shunting") the current to brain areas adjacent or distant to the brain region directly underneath a given electrode. Computational forward models of current flow provide estimates for the direction and magnitude of electric fields within the brain by accounting for the conductivity of anatomical features, thereby providing information that can inform stimulation protocols to boost the tDCS effects, to help inferences of tDCS outcomes, and to contribute to resolving problems with inconsistent effects reported across studies (Bikson et al., 2012).

Early studies, for instance, using simplified spherical head models (e.g., concentric spheres) of tDC flow answered questions regarding the relationship between current intensity and electrode size, such as highlighting the heterogeneity of the current profile on smaller electrodes (Miranda et al., 2006), and helped explain the relationship between stimulation efficacy and distance between electrodes, with greater distance increasing current impendence (Datta et al., 2008; Miranda et al., 2006). Because these models ignored anatomical differences, however, they only provided crude inferences. Later, more realistic models included detailed representations of human skull and brain (Wagner et al., 2007), with recent studies focusing on MRI derived images to construct a realistic head model. Fine

resolution models of gyri sulci profiles revealed that in sulci – where CSF fluid gathers – concentration of current were found (so-called current "hotspots"), reinforcing the point that the excitability changes underneath the electrode may not be uniformly excitatory or inhibitory (Salvador et al., 2010), and that current dose should be adjusted to account for the intensity in hotspots. For example, in one study, gains on a working memory task were accrued in participants who showed higher concentrations of current in targeted left prefrontal regions (Kim et al., 2014). Similarly, models of cellular effects mean that one can estimate the transmembrane potential within an electric field, and therefore predict the excitability changes that might occur in response to an administered current (Rahman et al., 2013).

Although very few studies have sought to directly link tDCS outcomes with model predictions to tests its utility, work developing of HD-tDCS present interesting findings. HD-tDCS was developed to test the assumption that by reducing the distance between the anode and cathode, one can focalise the neurophysiological effect of tDCS. The result was a multi-electrode array in which a central electrode is flanked by four electrodes of the opposite polarity that serve to restrict current flow from the central electrode to within a smaller, proximal region compared to tDCS. The focality of HD-tDCS was tested by measuring MEPs induced by transcranial electrical stimulation (tES), a form of non-invasive brain stimulation that administers a very high intensity current for several milliseconds to evoke action potentials (Edwards et al., 2013). The models predicted that HD-tES would evoke an MEP following stimulation of the primary motor cortex, but not in areas slightly anterior or posterior to this location. The model predictions were surprisingly borne out by the results. Moreover, however, individual models were generated based on MRI derived structural scans of each participant, and the models predicted that each participant would vary in the required dose needed to evoke an MEP. In a separate experiment, 600mV of HD-tES was administered to participants, and the resulting MEPs varied in magnitude across participants, and this variation was in line with the model predictions. For instance, those individuals who were predicted to be less sensitive to HD-tES administered at 600mV intensity did not show MEPs that significantly differed from baseline. In other – albeit preliminary – studies tDCS and HD-tDCS were directly compared in terms of effects on motor cortical excitability as measured in terms of TMS evoked MEPs. The results showed

that although HD-tDCS failed to increase MEP amplitudes after stimulation, unlike tDCS HD-tDCS evoked MEPs showed a peak increase 30mins post stimulation cessation with excitatory effects persisting for a further 2 hours (Edwards et al., 2013; Kuo et al., 2013; Villamar, Volz, Bikson, Datta, DaSilva & Fregni, 2013).

It is clear after reading the literature that current flow models are likely to become standard practice. Various (sometimes free[13]) software platforms already exist that provide a quick and easy means to model current flow with different electrode placements (Jung et al., 2013), and modelling software is now sold with popular tDCS devices (such as the Soterix devices). However, models have various important shortcomings that should be considered. At present, there is no standard procedure for generating a model, and the many decisions made at each stage of the model building process vary widely across studies, a problem exacerbated when constructing sophisticated models that attempt to account for the complexity of the human brain. For example, there is disagreement on what conductivity values should be assigned to key tissues such as bone, CSF, skin, fat, neural tissue, which mean that studies can depart from one another in their estimates of current flow in response to a given dose of current (for further discussion, see Bikson et al., 2012). More importantly, however, the conductivity of neural tissue will also depend on its excitability, which is liable to change when the brain is engaged in a cognitive task. However, at present, conductivity values are fixed, and generating a model that can dynamically adjusts conductivity values to accurately account for changes in brain activity presents a considerable challenge (see Bikson et al., 2012; Seo & Jun, 2017). This therefore limits the utility of current models especially for studies that measure tDCS effects on cognitive performance. A more basic concern, however is that with no means to view current flow in the brain, it is difficult to assess how realistic current models are. One way to validate models is to compare model predictions with study outcomes, but very few studies have done this, with research on HD-tDCS being the notable

[13] One database (www.neuralengr. com/bonsai) can be used to model current flow for popular two-electrode montages.

exception. However, although positive effects have been reported with HD-tDCS in learning new words (Perceval, Martin, Copland, Laine & Meinzer, 2017), working memory (Hussey et al., 2015; Nikolin, Loo, Bai, Dokos & Martin, 2015), and adaptive cognitive control (Gbadeyan, McMahon, Steinhauser & Meinzer, 2016), there is not enough to make a clear inference about its ability to modulate cognition. Thus, more research should be carried out to assess the effectiveness of HD-tDCS. Perhaps a more worrying concern for the future use of models, however, is that models may not be used to legitimately provide plausible accounts for research findings, but instead provide an expedient, post hoc interpretation for inconvenient findings, a possibility that should not be ignored given the findings reported in Section 7.2.1.

7.4 Conclusion

The null findings reported in this thesis are not an encouraging indicator for the effectiveness of tDCS, nor do they show in any good light the rigor in which investigations are carried out. However, I acknowledge that these null effects speak only to the conditions I investigated, namely the application of single sessions of tDCS on mainly frontal regions whilst participants performed word production and/or working memory tasks. More importantly, however, these null findings should not be taken as evidence that tDCS should be scrapped. Instead, they should act as a spur to encourage further but more rigorous investigations into the conditions in which tDCS can operate effectively, which are carried out with a more open approach that does not shy away from replication or reporting null/inconsistent effects. As previously stated in the close of Chapter 3: novel interventions typically pass through a hype cycle, where there is an initial spike in interest, which then diminishes as expectation of its efficacy are dealt repeated splashes of cold water from reports of null, inconsistent and small effects. What must not be forgotten, however, is that this period of disillusionment is superseded by a slope of enlightenment until we reach a *plateau of productivity*. The onus is, therefore, on future research to rise to this plateau by standing up to the full height of scientific rigour and merit.

References

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 604-614.

Alonzo, A., Brassil, J., Taylor, J. L., Martin, D., & Loo, C. K. (2012). Daily transcranial direct current stimulation (tDCS) leads to greater increases in cortical excitability than second daily transcranial direct current stimulation. *Brain stimulation, 5*(3), 208-213.

Amadi, U., Allman, C., Johansen-Berg, H., & Stagg, C. J. (2015). The homeostatic interaction between anodal transcranial direct current stimulation and motor learning in humans is related to GABA A activity. *Brain stimulation, 8*(5), 898-905.

Ambrus, G. G., Paulus, W., & Antal, A. (2010). Cutaneous perception thresholds of electrical stimulation methods: comparison of tDCS and tRNS. *Clinical Neurophysiology*, *121*(11), 1908-1914.

Ambrus, G. G., Zimmer, M., Kincses, Z. T., Harza, I., Kovács, G., Paulus, W., & Antal, A. (2011). The enhancement of cortical excitability over the DLPFC before and during training impairs categorization in the prototype distortion task. *Neuropsychologia*, *49*(7), 1974-1980.

Antal, A., Boros, K., Poreisz, C., Chaieb, L., Terney, D., & Paulus, W. (2008). Comparatively weak after-effects of transcranial alternating current stimulation (tACS) on cortical excitability in humans. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *1*(2), 97-105.

Antal, A., Keeser, D., Priori, A., Padberg, F., & Nitsche, M. A. (2015). Conceptual and procedural shortcomings of the systematic review "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" by Horvath and co-workers. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8*(4), 846-849.

Antal, A., Nitsche, M. A., Kincses, T. Z., Kruse, W., Hoffmann, K. P., & Paulus, W. (2004). Facilitation of visuo-motor learning by transcranial direct current stimulation of the motor and extrastriate visual areas in humans. *European Journal of Neuroscience, 19*(10), 2888-2892.

Antal, A., Polania, R., Schmidt-Samoa, C., Dechent, P., & Paulus, W. (2011). Transcranial direct current stimulation over the primary motor cortex during fMRI. *Neuroimage, 55*(2), 590-596.

Ardolino, G., Bossi, B., Barbieri, S., & Priori, A. (2005). Non-synaptic mechanisms underlie the after-effects of cathodal transcutaneous direct current stimulation of the human brain. *The Journal of physiology, 568*(2), 653-663.

Atkins, A. S., & Reuter-Lorenz, P. A. (2011). Neural mechanisms of semantic interference and false recognition in short-term memory. *NeuroImage, 56*(3), 1726-1734.

Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. A., Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory & cognition, 39(*5), 806-817.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). Distributed by the Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (2002). Is working memory still working?. *European psychologist*, *7*(2), 85.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders, 36*(3), 189-208.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in cognitive sciences, 12(*5), 193-200.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. Nature reviews. *Neuroscience, 10*(9), 659.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*(6), 907-918.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45*(13), 2883-2901.

Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology, 20(*5), 529-538.

Barak, O., & Tsodyks, M. (2014). Working models of working memory. *Current opinion in neurobiology, 25,* 20-24.

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: a meta-analytic study. *Applied neuropsychology, 15*(2), 97-106.

Bastani, A., & Jaberzadeh, S. (2013). Differential modulation of corticospinal excitability by different current densities of anodal transcranial direct current stimulation. *PLoS One, 8*(8), e72254.

Batsikadze, G., Moliadze, V., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Partially non-linear stimulation intensity-dependent effects of direct current stimulation on motor cortex excitability in humans. *The Journal of physiology, 591*(7), 1987-2000.

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review, 15*(2), 357-363.

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language, 69(*3), 228-256.

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *The Quarterly Journal of Experimental Psychology, 66*(11), 2135-2160.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia, 6*(1), 53-60.

Benwell, C. S., Learmonth, G., Miniussi, C., Harvey, M., & Thut, G. (2015). Non-linear effects of transcranial direct current stimulation as a function of individual baseline performance: evidence from biparietal tDCS influence on lateralized attention bias. C*ortex, 69*, 152-165.

Berberian, A. A., Moraes, G. V., Gadelha, A., Brietzke, E., Fonseca, A. O., Scarpato, B. S., ... & Lacerda, A. L. (2016). Is semantic verbal fluency impairment explained by executive function deficits in schizophrenia?. *Revista Brasileira de Psiquiatria, 38(*2), 121-126.

Berryhill, M. E., & Jones, K. T. (2012). tDCS selectively improves working memory in older adults with more education. *Neuroscience letters, 521*(2), 148-151.

Berryhill, M. E., & Jones, K. T. (2012). tDCS selectively improves working memory in older adults with more education. *Neuroscience letters, 521*(2), 148-151.

Berryhill, M. E., Peterson, D. J., Jones, K. T., & Stephens, J. A. (2014). Hits and misses: leveraging tDCS to advance cognitive research. *Frontiers in psychology*, *5.*

Besson, P., Perrey, S., Teo, W. P., & Muthalib, M. (2016). Commentary: Cumulative effects of anodal and priming cathodal tDCS on pegboard test performance and motor cortical excitability. *Frontiers in human neuroscience*, *10*.

Bestmann, S., de Berker, A. O., & Bonaiuto, J. (2015). Understanding the behavioural consequences of noninvasive brain stimulation. *Trends in cognitive sciences, 19*(1), 13-20.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. Bilingualism: *Language and cognition, 12*(1), 3-11.

Biesbroek, J. M., van Zandvoort, M. J., Kappelle, L. J., Velthuis, B. K., Biessels, G. J., & Postma, A. (2016). Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. *Brain Structure and Function*, *221*(4), 2123-2134.

Bikson, M., & Rahman, A. (2013). Origins of specificity during tDCS: anatomical, activity-selective, and input-bias mechanisms. *Frontiers in human neuroscience*, *7*, 688.

Bikson, M., Rahman, A., & Datta, A. (2012). Computational models of transcranial direct current stimulation. *Clinical EEG and Neuroscience*, *43*(3), 176-183.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences, 15*(11), 527-536.

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

Bindman, L. J., Lippold, O. C. J., & Redfearn, J. W. T. (1964). The action of brief polarizing currents on the cerebral cortex of the rat (1) during current flow and (2) in the production of long-lasting after-effects. *The Journal of physiology, 172(*3), 369-382.

Binney, R. J., Zuckerman, B. M., Waller, H. N., Hung, J., Ashaie, S. A., & Reilly, J. (2018). Cathodal tDCS of the Bilateral Anterior Temporal Lobes Facilitates Semantically-Driven Verbal Fluency. *Neuropsychologia*.

Bishop, N. A., Lu, T., & Yankner, B. A. (2010). Neural mechanisms of ageing and cognitive decline. *Nature*, *464*(7288), 529-535.

Bloom, J. S., & Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition?. *Neuropsychology review, 15(*2), 59-71.

Boehringer, A., Macher, K., Dukart, J., Villringer, A., & Pleger, B. (2013). Cerebellar transcranial direct current stimulation modulates verbal working memory. *Brain stimulation, 6*(4), 649-653.

Boggio, P. S., Fregni, F., Valasek, C., Ellwood, S., Chi, R., Gallate, J., ... & Snyder, A. (2009). Temporal lobe cortical electrical stimulation during the encoding and retrieval phase reduces false memories. *PloS one, 4*(3), e4959.

Boggio, P. S., Liguori, P., Sultani, N., Rezende, L., Fecteau, S., & Fregni, F. (2009). Cumulative priming effects of cortical stimulation on smoking cue-induced craving. *Neuroscience letters*, *463*(1), 82-86.

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2009. Introduction to Meta-analysis. John Wiley & Sons, West Sussex, United Kingdom.

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.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience, 14*(8), 585-585.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376.

Canini, M., Della Rosa, P. A., Catricalà, E., Strijkers, K., Branzi, F. M., Costa, A., & Abutalebi, J. (2016). Semantic interference and its control: A functional neuroimaging and connectivity study. *Human brain mapping, 37*(11), 4179-4196.

Cappon, D., Jahanshahi, M., & Bisiacchi, P. (2016). Value and efficacy of transcranial direct current stimulation in the cognitive rehabilitation: a critical review since 2000. *Frontiers in neuroscience, 10*.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of personality and social psychology, 67(*2), 319-333.

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.

Cattaneo, Z., Pisoni, A., Gallucci, M., & Papagno, C. (2016). tDCS effects on verbal fluency: A response to Vannorsdall et al (2016). *Cognitive and Behavioral Neurology, 29*(3), 117-121.

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.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*(1), 40-48.

Chaieb, L., Antal, A., & Paulus, W. (2011). Transcranial alternating current stimulation in the low kHz range increases motor cortex excitability. *Restorative neurology and neuroscience*, *29*(3), 167-175.

Chaieb, L., Antal, A., & Paulus, W. (2015). Transcranial random noise stimulation-induced plasticity is NMDA-receptor independent but sodium-channel blocker and benzodiazepines sensitive. *Frontiers in neuroscience*, *9*.

Chaieb, L., Kovacs, G., Cziraki, C., Greenlee, M., Paulus, W., & Antal, A. (2009). Short-duration transcranial random noise stimulation induces blood oxygenation level dependent response attenuation in the human motor cortex. *Experimental brain research*, *198*(4), 439-444.

Chesters, J., Watkins, K. E., & Möttönen, R. (2017). Investigating the feasibility of using transcranial direct current stimulation to enhance fluency in people who stutter. *Brain and language, 164*, 68-76.

Chesters, J., Watkins, K. E., & Möttönen, R. (2017). Investigating the feasibility of using transcranial direct current stimulation to enhance fluency in people who stutter. *Brain and language*, *164*, 68-76.

Chew, T., Ho, K. A., & Loo, C. K. (2015). Inter-and intra-individual variability in response to transcranial direct current stimulation (tDCS) at varying current intensities. *Brain Stimulation, 8*(6), 1130-1137.

Chiarello, C., & Maxfield, L. (1996). Varieties of interhemispheric inhibition, or how to keep a good hemisphere down. *Brain and cognition, 30(*1), 81-108.

Chrysikou, E. G., & Hamilton, R. H. (2011). Noninvasive brain stimulation in the treatment of aphasia: exploring interhemispheric relationships and their implications for neurorehabilitation. *Restorative neurology and neuroscience, 29*(6), 375-394.

Clark, V. P., Coffman, B. A., Trumbo, M. C., & Gasparovic, C. (2011). Transcranial direct current stimulation (tDCS) produces localized and specific alterations in neurochemistry: a 1 H magnetic resonance spectroscopy study. *Neuroscience letters, 500*(1), 67-71.

Coffman, B. A., Clark, V. P., & Parasuraman, R. (2014). Battery powered thought: enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *Neuroimage*, *85*, 895-908.

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why?. *Intelligence, 36*(6), 584-606.

Cools, R., & D'Esposito, M. (2011). Inverted-U–shaped dopamine actions on human working memory and cognitive control. *Biological psychiatry, 69(*12), e113-e125.

Costanzo, F., Varuzza, C., Rossi, S., Sdoia, S., Varvara, P., Oliveri, M., ... & Menghini, D. (2016). Evidence for reading improvement following tDCS treatment in children and adolescents with Dyslexia. *Restorative neurology and neuroscience, 34*(2), 215-226.

Crinion, J. T. (2016). Transcranial direct current stimulation as a novel method for enhancing aphasia treatment effects. European Psychologist.

Cumming, G. (2013). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.

Cuypers, K., Leenus, D. J., van den Berg, F. E., Nitsche, M. A., Thijs, H., Wenderoth, N., & Meesen, R. L. (2013). Is motor learning mediated by tDCS intensity?. *PLoS One, 8(*6), e67344.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1-2), 179-229.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior, 19(*4), 450-466.

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.

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.

de Aguiar, V., Paolazzi, C. L., & Miceli, G. (2015). tDCS in post-stroke aphasia: the role of stimulation parameters, behavioral treatment and patient characteristics. *Cortex, 63*, 296-316.

De Berker, A. O., Bikson, M., & Bestmann, S. (2013). Predicting the behavioral impact of transcranial direct current stimulation: issues and limitations. *Frontiers in human neuroscience, 7.*

De Vries, M. H., Barth, A. C., Maiworm, S., Knecht, S., Zwitserlood, P., & Flöel, A. (2010). Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar.*Journal of cognitive neuroscience*, *22*(11), 2427-2436.

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.

Dedoncker, J., Brunoni, A. R., Baeken, C., & Vanderhasselt, M. A. (2016). The effect of the interval-between-sessions on prefrontal transcranial direct current stimulation (tDCS) on cognitive outcomes: a systematic review and meta-analysis. *Journal of Neural Transmission, 123*(10), 1159-1172.

Devlin, J. T., & Watkins, K. E. (2006). Stimulating language: insights from TMS. *Brain, 130*(3), 610-622.

Díez, E., Gómez-Ariza, C. J., Díez-Álamo, A. M., Alonso, M. A., & Fernandez, A. (2017). Encoding/retrieval dissociation of false recognition with transcranial direct current stimulation (tDCS) of the left temporal lobe. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 10(*2), 372.

Dockery, C. A., Hueckel-Weng, R., Birbaumer, N., & Plewnia, C. (2009). Enhancement of planning ability by transcranial direct current stimulation. *Journal of Neuroscience, 29*(22), 7271-7277.

Dondé, C., Amad, A., Nieto, I., Brunoni, A. R., Neufeld, N. H., Bellivier, F., ... & Geoffroy, P. A. (2017). Transcranial direct-current stimulation (tDCS) for bipolar depression: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*.

Donner, T. H., & Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends in cognitive sciences*, *15*(5), 191-199.

Dresler, M., Sandberg, A., Ohla, K., Bublitz, C., Trenado, C., Mroczko-Wąsowicz, A., ... & Repantis, D. (2013). Non-pharmacological cognitive enhancement. *Neuropharmacology*, *64*, 529-543.

Edwards, D., Cortes, M., Datta, A., Minhas, P., Wassermann, E. M., & Bikson, M. (2013). Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: a basis for high-definition tDCS. *Neuroimage*, *74*, 266-275.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj, 315*(7109), 629-634.

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*(1), 29.

Elsner, B., Kugler, J., Pohl, M., & Mehrholz, J. (2015). Transcranial direct current stimulation (tDCS) for improving aphasia in patients with aphasia after stroke. The Cochrane Library.

Emmerling, F., Martijn, C., Alberts, H. J., Thomson, A. C., David, B., Kessler, D., ... & Sack, A. T. (2017). The (non-) replicability of regulatory resource depletion: A field report employing non-invasive brain stimulation. *PloS one, 12*(3), e0174331.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, *2*(10), 704.

Esmaeilpour, Z., Marangolo, P., Hampstead, B. M., Bestmann, S., Galletta, E., Knotkova, H., & Bikson, M. (2018). Incomplete evidence that increasing current intensity of tDCS boosts outcomes. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *11*(2), 310-321.

Feredoes, E., Tononi, G., & Postle, B. R. (2006). Direct evidence for a prefrontal contribution to the control of proactive interference in verbal working memory. *Proceedings of the National Academy of Sciences, 103*(51), 19530-19534.

Ferrucci, R., Mameli, F., Guidi, I., Mrakic-Sposta, S., Vergari, M., Marceglia, S. E. E. A., ... & Priori, A. (2008). Transcranial direct current stimulation improves recognition memory in Alzheimer disease. *Neurology, 71*(7), 493-498.

Ferrucci, R., Marceglia, S., Vergari, M., Cogiamanian, F., Mrakic-Sposta, S., Mameli, F. E. E. A., ... & Priori, A. (2008). Cerebellar transcranial direct current stimulation impairs the practice-dependent proficiency increase in working memory. *Journal of Cognitive Neuroscience, 20*(9), 1687-1697.

Fertonani, A., & Miniussi, C. (2017). Transcranial electrical stimulation: what we know and do not know about mechanisms. *The Neuroscientist, 23*(2), 109-123.

Fertonani, A., Brambilla, M., Cotelli, M., & Miniussi, C. (2014). The timing of cognitive plasticity in physiological aging: a tDCS study of naming. *Frontiers in aging neuroscience, 6*.

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.

Fertonani, A., Pirulli, C., & Miniussi, C. (2011). Random noise stimulation improves neuroplasticity in perceptual learning. *Journal of Neuroscience*, *31*(43), 15416-15423.

Feurra, M., Bianco, G., Santarnecchi, E., Del Testa, M., Rossi, A., & Rossi, S. (2011). Frequency-dependent tuning of the human motor system induced by transcranial oscillatory potentials. *Journal of Neuroscience*, *31*(34), 12165-12170.

Fiori, V., Coccia, M., Marinelli, C. V., Vecchi, V., Bonifazi, S., Ceravolo, M. G., ... & Marangolo, P. (2011). Transcranial direct current stimulation improves word

retrieval in healthy and nonfluent aphasic subjects. *Journal of Cognitive Neuroscience, 23(*9), 2309-2323.

Fino, E., Melogno, S., Iliceto, P., D'Aliesio, S., Pinto, M. A., Candilera, G., & Sabatello, U. (2014). Executive functions, impulsivity, and inhibitory control in adolescents: A structural equation model. *Advances in cognitive psychology*, *10*(2), 32.

Flöel, A., Meinzer, M., Kirstein, R., Nijhof, S., Deppe, M., Knecht, S., & Breitenstein, C. (2011). Short-term anomia training and electrical brain stimulation. *Stroke, 42*(7), 2065-2067.

Flöel, A., Rösser, N., Michka, O., Knecht, S., & Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. *Journal of Cognitive Neuroscience, 20*(8), 1415-1422.

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2016). Detecting and avoiding likely false-positive findings–a practical guide. *Biological Reviews*.

Francis, J. T., Gluckman, B. J., & Schiff, S. J. (2003). Sensitivity of neurons to weak electric fields. *Journal of Neuroscience*, *23*(19), 7255-7261.

Fricke, K., Seeber, A. A., Thirugnanasambandam, N., Paulus, W., Nitsche, M. A., & Rothwell, J. C. (2011). Time course of the induction of homeostatic plasticity generated by repeated transcranial direct current stimulation of the human motor cortex.*Journal of neurophysiology*, *105*(3), 1141-1149.

Furuya, S., Klaus, M., Nitsche, M. A., Paulus, W., & Altenmüller, E. (2014). Ceiling effects prevent further improvement of transcranial stimulation in skilled musicians. *Journal of Neuroscience, 34*(41), 13834-13839.

Gallasch, E., Christova, M., Kunz, A., Rafolt, D., & Golaszewski, S. (2015). Modulation of sensorimotor cortex by repetitive peripheral magnetic stimulation. *Frontiers in human neuroscience,9*.

Gathercole, S. E., & Baddeley, A. D. (2014). Working memory and language. Psychology Press.

Gauvin, H. S., Meinzer, M., & de Zubicaray, G. I. (2017). tDCS effects on word production: Limited by design? Comment on Westwood et al.(2017). Cortex; a journal devoted to the study of the nervous system and behavior.

Gbadeyan, O., McMahon, K., Steinhauser, M., & Meinzer, M. (2016). Stimulation of dorsolateral prefrontal cortex enhances adaptive cognitive control: a high-definition

transcranial direct current stimulation study. *Journal of Neuroscience*, *36*(50), 12530-12536.

Giordano, J., Bikson, M., Kappenman, E. S., Clark, V. P., Coslett, H. B., Hamblin, M. R., ... & Nitsche, M. A. (2017). Mechanisms and effects of transcranial direct current stimulation. *Dose-Response, 15*(1), 1559325816685467.

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.

Gözenman, F., & Berryhill, M. E. (2016). Working memory capacity differentially influences responses to tDCS and HD-tDCS in a retro-cue task. *Neuroscience letters, 629,* 105-109.

Gray, J. R., & Braver, T. S. (2002). Personality predicts working-memory—related activation in the caudal anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(1), 64-75.

Gray, J. R., & Burgess, G. C. (2004). Personality differences in cognitive control? BAS, processing efficiency, and the prefrontal cortex. *Journal of Research in Personality*, *38*(1), 35-36.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature neuroscience*, *6*(3), 316.

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience, 5*(1), 1-13.

Hamilton, A. C., & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43*(1), 112-123.

Hamilton, R. H., Chrysikou, E. G., & Coslett, B. (2011). Mechanisms of aphasia recovery after stroke and the role of noninvasive brain stimulation. *Brain and language, 118(*1), 40-50.

Harris, R. (2017). *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. Hachette UK.

Harris, S. (2014). *Waking up: A guide to spirituality without religion*. Simon and Schuster.

Hartwigsen, G., Price, C. J., Baumgaertner, A., Geiss, G., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). The right posterior inferior frontal gyrus contributes to

phonological word decisions in the healthy brain: evidence from dual-site TMS. *Neuropsychologia, 48*(10), 3155-3163.

Henseler, I., Mädebach, A., Kotz, S. A., & Jescheniak, J. D. (2014). Modulating brain mechanisms resolving lexico-semantic Interference during word production: A transcranial direct current stimulation study. *Journal of cognitive neuroscience, 26*(7), 1403-1417.

Herrmann, C. S., Rach, S., Neuling, T., & Strüber, D. (2013). Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Frontiers in human neuroscience*, *7*.

Heth, I., & Lavidor, M. (2015). Improved reading measures in adults with dyslexia following transcranial direct current stimulation treatment. *Neuropsychologia, 70,* 107-113.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407-422.

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.

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. *Neuropsychologia, 44*(12), 2547-2557.

Hitchens, P. (2016). *War We Never Fought*. Featherstone Education.

Holland, R., & Crinion, J. (2012). Can tDCS enhance treatment of aphasia after stroke?. *Aphasiology, 26*(9), 1169-1191.

Holland, R., Leff, A. P., Josephs, O., Galea, J. M., Desikan, M., Price, C. J., ... & Crinion, J. (2011). Speech facilitation by left inferior frontal cortex stimulation. *Current Biology, 21*(16), 1403-1407.

Holland, R., Leff, A. P., Penny, W. D., Rothwell, J. C., & Crinion, J. (2016). Modulation of frontal effective connectivity during speech. *NeuroImage, 140*, 126-133.

Hopfinger, J. B., Parsons, J., & Fröhlich, F. (2017). Differential effects of 10-Hz and 40-Hz transcranial alternating current stimulation (tACS) on endogenous versus exogenous attention. *Cognitive neuroscience*, *8*(2), 102-111.

Horvath, J. C. (2015). Are current blinding methods for transcranial direct current stimulation (tDCS) effective in healthy populations?. *Clinical Neurophysiology*, *126*(11), 2045-2046.

Horvath, J. C. (2015). New Quantitative Analyses Following Price & Hamilton's Critique do not Change Original Findings of Horvath et al. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8*(3), 665-666.

Horvath, J. C., Carter, O., & Forte, J. D. (2014). Transcranial direct current stimulation: five important issues we aren't discussing (but probably should be). *Frontiers in systems neuroscience, 8*.

Horvath, J. C., Carter, O., & Forte, J. D. (2016). No significant effect of transcranial direct current stimulation (tDCS) found on simple motor reaction time comparing 15 different simulation protocols. *Neuropsychologia, 91*, 544-552.

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.

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.

Horvath, J. C., Vogrin, S. J., Carter, O., Cook, M. J., & Forte, J. D. (2016). Effects of a common transcranial direct current stimulation (tDCS) protocol on motor evoked potentials found to be highly variable within individuals over 9 testing sessions. *Experimental brain research, 234*(9), 2629-2642.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464-482.

Hoy, K. E., Emonson, M. R., 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.

Hsu, T. Y., Juan, C. H., & Tseng, P. (2016). Individual differences and state-dependent responses in transcranial direct current stimulation. *Frontiers in human neuroscience, 10*.

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.

Hussey, E. K., Ward, N., Christianson, K., & Kramer, A. F. (2015). Language and memory improvements following tDCS of left lateral prefrontal cortex. *PloS one*, *10*(11), e0141417.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology, 2*.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2*(8), e124.

Irlbacher, K., Kraft, A., Kehrer, S., & Brandt, S. A. (2014). Mechanisms and neuronal networks involved in reactive and proactive cognitive control of interference in working memory. *Neuroscience & Biobehavioral Reviews, 46*, 58-70.

Iyer, M. B., Mattu, U., Grafman, J., Lomarev, M., Sato, S., & Wassermann, E. M. (2005). Safety and cognitive effect of frontal DC brain polarization in healthy individuals. *Neurology, 64*(5), 872-875.

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Experimental brain research, 216*(1), 1-10.

Jalali, R., Miall, R. C., & Galea, J. M. (2017). No consistent effect of cerebellar transcranial direct current stimulation on visuomotor adaptation. *Journal of Neurophysiology, 118*(2), 655-665.

Javadi, A. H., & Cheng, P. (2013). Transcranial direct current stimulation (tDCS) enhances reconsolidation of long-term memory. *Brain stimulation*, *6*(4), 668-674.

Javadi, A. H., & Walsh, V. (2012). Transcranial direct current stimulation (tDCS) of the left dorsolateral prefrontal cortex modulates declarative memory. *Brain stimulation*, *5*(3), 231-241.

Javadi, A. H., Cheng, P., & Walsh, V. (2012). Short duration transcranial direct current stimulation (tDCS) modulates verbal memory. *Brain stimulation*, *5*(4), 468-474.

Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex, 49*(3), 611-625.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain, 129*(8), 2132-2147.

Jefferson, S., Mistry, S., Singh, S., Rothwell, J., & Hamdy, S. (2009). Characterizing the application of transcranial direct current stimulation in human pharyngeal motor cortex. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 297*(6), G1035-G1040.

Jeon, S. Y., & Han, S. J. (2012). Improvement of the working memory and naming by transcranial direct current stimulation. *Annals of rehabilitation medicine, 36*(5), 585-595.

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., ... & Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology, 27*(2), 220.

Jones, K. T., & Berryhill, M. E. (2012). Parietal contributions to visual working memory depend on task difficulty. *Frontiers in psychiatry, 3*.

Jones, K. T., Gözenman, 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.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139*(1), 181-193.

Joseph, A., 2011. Plot Digitizer 2.5.1.

Joundi, R. A., Jenkinson, N., Brittain, J. S., Aziz, T. Z., & Brown, P. (2012). Driving oscillatory activity in the human cortex enhances motor performance. *Current Biology*, *22*(5), 403-407.

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., ... & Diefenbach, P. (2017). No Effect of Commercial Cognitive Training on Brain Activity, Choice Behavior, or Cognitive Performance. *Journal of Neuroscience, 37*(31), 7390-7402.

Kadosh, R. C., Soskic, S., Iuculano, T., Kanai, R., & Walsh, V. (2010). Modulating neuronal activity produces specific and long-lasting changes in numerical competence. *Current Biology, 20*(22), 2016-2020.

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 43-57.

Kang, E. K., Kim, Y. K., Sohn, H. M., Cohen, L. G., & Paik, N. J. (2011). Improved picture naming in aphasia patients treated with cathodal tDCS to inhibit the right Broca's homologue area. *Restorative neurology and neuroscience, 29*(3), 141-152.

Kessler, Y., & Meiran, N. (2008). Two dissociable updating processes in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1339.

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic bulletin & review*, *22*(6), 1770-1776.

Kasten, F. H., Dowsett, J., & Herrmann, C. S. (2016). Sustained aftereffect of α-tACS lasts up to 70 min after stimulation. *Frontiers in human neuroscience*, *10*, 245.

Kidgell, D. J., Daly, R. M., Young, K., Lum, J., Tooley, G., Jaberzadeh, S., ... & Pearce, A. J. (2013). Different current intensities of anodal transcranial direct current stimulation do not differentially modulate motor cortex plasticity. *Neural plasticity,* 2013.

Kim, J. H., Kim, D. W., Chang, W. H., Kim, Y. H., Kim, K., & Im, C. H. (2014). Inconsistent outcomes of transcranial direct current stimulation may originate from anatomical differences among individuals: electric field simulation using individual MRI data. *Neuroscience letters, 564*, 6-10.

Kiyonaga, A., & Egner, T. (2014). The working memory Stroop effect: when internal representations clash with external stimuli. *Psychological science, 25*(8), 1619-1629.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability. *Social psychology*.

Kohls, N., Sauer, S., Offenbächer, M., & Giordano, J. (2011). Spirituality: an overlooked predictor of placebo effects?. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1572), 1838-1848.

Krause, B., & Kadosh, R. C. (2014). Not all brains are created equal: the relevance of individual differences in responsiveness to transcranial electrical stimulation. *Frontiers in systems neuroscience, 8*.

Krause, B., Márquez-Ruiz, J., & Kadosh, R. C. (2013). The effect of transcranial direct current stimulation: a role for cortical excitation/inhibition balance?. *Frontiers in human neuroscience, 7*.

Kuo, H. I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Comparing cortical plasticity induced by conventional and high-definition 4× 1 ring tDCS: a neurophysiological study. *Brain stimulation*, *6*(4), 644-648.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44(*4), 978-990.

Kuo, H. I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Comparing cortical plasticity induced by conventional and high-definition 4× 1 ring tDCS: a neurophysiological study. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *6*(4), 644-648.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*.

Lambon, R. M., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience, 13*(3), 341-356.

Lang, N., Siebner, H. R., Ward, N. S., Lee, L., Nitsche, M. A., Paulus, W., ... & Frackowiak, R. S. (2005). How does transcranial DC stimulation of the primary motor cortex alter regional neuronal activity in the human brain?. *European Journal of Neuroscience, 22*(2), 495-504.

Lapenta, O. M., Minati, L., Fregni, F., & Boggio, P. S. (2013). Je pense donc je fais: transcranial direct current stimulation modulates brain oscillations associated with motor imagery and movement observation. *Frontiers in human neuroscience, 7*.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920.

Lawrence, B. J., Gasson, N., Bucks, R. S., Troeung, L., & Loftus, A. M. (2017). Cognitive Training and Noninvasive Brain Stimulation for Cognition in Parkinson's Disease: A Meta-analysis. *Neurorehabilitation and Neural Repair*, 1545968317712468.

Lazar, R. M., & Mohr, J. P. (2011). Revisiting the contributions of Paul Broca to the study of aphasia. *Neuropsychology review*, *21*(3), 236.

Learmonth, G., Thut, G., Benwell, C. S., & Harvey, M. (2015). The implications of state-dependent tDCS effects in aging: behavioural response is determined by baseline performance. *Neuropsychologia, 74*, 108-119.

Lee, S. Y., Cheon, H. J., Yoon, K. J., Chang, W. H., & Kim, Y. H. (2013). Effects of dual transcranial direct current stimulation for aphasia in chronic stroke patients. *Annals of rehabilitation medicine, 37(*5), 603-610.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1-22.

Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences*, *3*(6), 223-232.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38.

Li, L. M., Uehara, K., & Hanakawa, T. (2015). The contribution of interindividual factors to variability of response in transcranial direct current stimulation studies. *Frontiers in cellular neuroscience, 9.*

Liebetanz, D., Koch, R., Mayenfels, S., König, F., Paulus, W., & Nitsche, M. A. (2009). Safety limits of cathodal transcranial direct current stimulation in rats. *Clinical Neurophysiology, 120*(6), 1161-1167.

Liebetanz, D., Nitsche, M. A., Tergau, F., & Paulus, W. (2002). Pharmacological approach to the mechanisms of transcranial DC-stimulation-induced after-effects of human motor cortex excitability. *Brain, 125*(10), 2238-2247.

Lippold, O. C. J., & Redfearn, J. W. T. (1964). Mental changes resulting from the passage of small direct currents through the human brain. *The British Journal of Psychiatry, 110(*469), 768-772.

Liuzzi, G., Freundlieb, N., Ridder, V., Hoppe, J., Heise, K., Zimerman, M., ... & Hummel, F. C. (2010). The involvement of the left motor cortex in learning of a novel action word lexicon.*Current Biology*, *20*(19), 1745-1751.

London, R. E., & Slagter, H. A. (2015). Effects of transcranial direct current stimulation over left dorsolateral pFC on the attentional blink depend on individual baseline performance. *Journal of cognitive neuroscience, 27*(12), 2382-2393.

Looi, C. Y., Duta, M., Brem, A. K., Huber, S., Nuerk, H. C., & Kadosh, R. C. (2016). Combining brain stimulation and video game to promote long-term transfer of learning and cognitive enhancement. *Scientific reports, 6*, 22003.

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., & Fernández-del-Olmo, M. (2014). Inter-individual variability in response to non-invasive brain stimulation paradigms. *Brain stimulation, 7*(3), 372-380.

López-Alonso, V., Fernández-del-Olmo, M., Costantini, A., Gonzalez-Henriquez, J. J., & Cheeran, B. (2015). Intra-individual variability in the response to anodal transcranial direct current stimulation. *Clinical Neurophysiology, 126*(12), 2342-2347.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 503.

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*, *28*(8), 1063-1089.

Manenti, R., Brambilla, M., Petesi, M., Ferrari, C., & Cotelli, M. (2013). Enhancing verbal episodic memory in older and young subjects after non-invasive brain stimulation. *Frontiers in aging neuroscience, 5*.

Marangolo, P., Fiori, V., Calpagnano, M. A., Campana, S., Razzano, C., Caltagirone, C., & Marini, A. (2013). tDCS over the left inferior frontal cortex improves speech production in aphasia. *Frontiers in Human Neuroscience, 7*.

Mark W.. Lipsey, & Wilson, D. B. (2001). Practical meta-analysis (Vol. 49). Thousand Oaks, CA: Sage publications.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., ... & Barone, P. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, *522*(1), 225-259.

Marshall, L., Mölle, M., Siebner, H. R., & Born, J. (2005). Bifrontal transcranial direct current stimulation slows reaction time in a working memory task. *BMC neuroscience, 6*(1), 23.

Martin, A., Meinzer, M., Lindenberg, R., Sieg, M. M., Nachtigall, L., & Flöel, A. (2017). Effects of Transcranial Direct Current Stimulation on Neural Network Structure in Young and Older Adults. *Journal of Cognitive Neuroscience.*

Martin, D. M., Liu, R., Alonzo, A., Green, M., & Loo, C. K. (2014). Use of transcranial direct current stimulation (tDCS) to enhance cognitive training: effect of timing of stimulation. *Experimental brain research, 232*(10), 3345-3351.

Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology, 8*.

Matsumoto, H., & Ugawa, Y. (2017). Adverse events of tDCS and tACS: a review. *Clinical Neurophysiology Practice, 2*, 19-25.

Matzen, L. E., Trumbo, M. C., Leach, R. C., & Leshikar, E. D. (2015). Effects of non-invasive brain stimulation on associative memory. *Brain research*, *1624*, 286-296.

Medeiros, Liciane Fernandes, Izabel Cristina Custodio de Souza, Liliane Pinto Vidor, Andressa de Souza, Alícia Deitos, Magdalena Sarah Volz, Felipe Fregni, Wolnei Caumo, and Iraci LS Torres. "Neurobiological effects of transcranial direct current stimulation: a review." *Frontiers in psychiatry 3* (2012).

Medina, J., & Cason, S. (2017). No evidential value in samples of transcranial direct current stimulation (tDCS) studies of cognition and working memory in healthy populations. *Cortex, 94*, 131-141.

Meinzer, M., Antonenko, D., Lindenberg, R., Hetzer, S., Ulm, L., Avirame, K., ... & Flöel, A. (2012). Electrical brain stimulation improves cognitive performance by modulating functional connectivity and task-specific activation. *Journal of Neuroscience, 32*(5), 1859-1866.

Meinzer, M., Jähnigen, S., Copland, D. A., Darkow, R., Grittner, U., Avirame, K., ... & Flöel, A. (2014). Transcranial direct current stimulation over multiple days improves learning and maintenance of a novel vocabulary. *Cortex, 50*, 137-147.

Meinzer, M., Lindenberg, R., Darkow, R., Ulm, L., Copland, D., & Flöel, A. (2014a). Transcranial direct current stimulation and simultaneous functional magnetic resonance imaging. *Journal of visualized experiments: JoVE*, (86).

Meinzer, M., Lindenberg, R., Sieg, M. M., Nachtigall, L., Ulm, L., & Flöel, A. (2014b). Transcranial direct current stimulation of the primary motor cortex improves word-retrieval in older adults. *Frontiers in aging neuroscience, 6*.

Meinzer, M., Yetim, Ö., McMahon, K., & de Zubicaray, G. (2016). Brain mechanisms of semantic interference in spoken word production: An anodal transcranial Direct Current Stimulation (atDCS) study. *Brain and language, 157*, 72-80.

Merrill, D. R., Bikson, M., & Jefferys, J. G. (2005). Electrical stimulation of excitable tissue: design of efficacious and safe protocols. *Journal of neuroscience methods, 141(*2), 171-198.

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: Basic, Translational, and Clinical Research in Neuromodulation*, *5*(2), 110-115.

Minarik, T., Berger, B., Althaus, L., Bader, V., Biebl, B., Brotzeller, F., ... & Leitner, M. (2016). The importance of sample size for reproducibility of tDCS effects. *Frontiers in human neuroscience, 10*.

Miniussi, C., Harris, J. A., & Ruzzoli, M. (2013). Modelling non-invasive brain stimulation in cognitive neuroscience. *Neuroscience & Biobehavioral Reviews, 37(*8), 1702-1712.

Miniussi, C., Ruzzoli, M., & Walsh, V. (2010). The mechanism of transcranial magnetic stimulation in cognition. *Cortex, 46*(1), 128-130.

Miranda, P. C., Lomarev, M., & Hallett, M. (2006). Modeling the current distribution during transcranial direct current stimulation. *Clinical neurophysiology, 117*(7), 1623-1629.

Mirman, D., & Britt, A. E. (2014). What we talk about when we talk about access deficits. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1634), 20120388.

Moliadze, V., Antal, A., & Paulus, W. (2010). Electrode-distance dependent after-effects of transcranial direct and random noise stimulation with extracephalic reference electrodes. *Clinical Neurophysiology*, *121*(12), 2165-2171.

Moliadze, V., Atalay, D., Antal, A., & Paulus, W. (2012). Close to threshold transcranial electrical stimulation preferentially activates inhibitory networks before switching to excitation with higher intensities. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *5*(4), 505-511.

Monte-Silva, K., Kuo, M. F., Hessenthaler, S., Fresnoza, S., Liebetanz, D., Paulus, W., & Nitsche, M. A. (2013). Induction of late LTP-like plasticity in the human motor cortex by repeated non-invasive brain stimulation. *Brain stimulation, 6(*3), 424-432.

Monte-Silva, K., Liebetanz, D., Grundey, J., Paulus, W., & Nitsche, M. A. (2010). Dosage-dependent non-linear effect of l-dopa on human motor cortex plasticity. *The Journal of physiology, 588*(18), 3415-3424.

Monti, A., Cogiamanian, F., Marceglia, S., Ferrucci, R., Mameli, F., Mrakic-Sposta, S., ... & Priori, A. (2008). Improved naming after transcranial direct current stimulation in aphasia. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(4), 451-453.

Monti, A., Ferrucci, R., Fumagalli, M., Mameli, F., Cogiamanian, F., Ardolino, G., & Priori, A. (2013). Transcranial direct current stimulation (tDCS) and language. *J Neurol Neurosurg Psychiatry*, *84*(8), 832-842.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*(11), 1723-1735.

Mulquiney, P. G., Hoy, K. E., Daskalakis, Z. J., & Fitzgerald, P. B. (2011). Improving working memory: exploring the effect of transcranial random noise stimulation and transcranial direct current stimulation on the dorsolateral prefrontal cortex. *Clinical Neurophysiology, 122*(12), 2384-2389.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021.

Murphy, T. H., & Corbett, D. (2009). Plasticity during stroke recovery: from synapse to behaviour. *Nature reviews. Neuroscience*, *10*(12), 861.

Nagel, I. E., Schumacher, E. H., Goebel, R., & D'esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *Neuroimage*, *43*(4), 801-807.

Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta psychologica*, *134*(3), 279-289.

Nelson, J. K., Reuter-Lorenz, P. A., Persson, J., Sylvester, C. Y. C., & Jonides, J. (2009). Mapping interference resolution across task domains: A shared control process in left inferior frontal gyrus. *Brain research, 1256*, 92-100.

Nikolin, S., Loo, C. K., Bai, S., Dokos, S., & Martin, D. M. (2015). Focalised stimulation using high definition transcranial direct current stimulation (HD-tDCS) to

investigate declarative verbal learning and memory functioning. *Neuroimage*, *117*, 11-19.

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*, 0956797617698139.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of physiology*, *527*(3), 633-639.

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology, 57(*10), 1899-1901.

Nitsche, M. A., Bikson, M., & Bestmann, S. (2015). On the use of meta-analysis in neuromodulatory non-invasive brain stimulation. *Brain Stimulation*, *8*(3), 66.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., ... & Pascual-Leone, A. (2008). Transcranial direct current stimulation: state of the art 2008. *Brain stimulation, 1*(3), 206-223.

Nitsche, M. A., Doemkes, S., Karakoese, T., Antal, A., Liebetanz, D., Lang, N., ... & Paulus, W. (2007). Shaping the effects of transcranial direct current stimulation of the human motor cortex. *Journal of neurophysiology, 97*(4), 3109-3117.

Nitsche, M. A., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., ... & Paulus, W. (2003). Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. *The Journal of physiology, 553*(1), 293-301.

Nitsche, M. A., Grundey, J., Liebetanz, D., Lang, N., Tergau, F., & Paulus, W. (2004). Catecholaminergic consolidation of motor cortical neuroplasticity in humans. *Cerebral Cortex, 14*(11), 1240-1245.

Nitsche, M. A., Jaussi, W., Liebetanz, D., Lang, N., Tergau, F., & Paulus, W. (2004). Consolidation of human motor cortical neuroplasticity by D-cycloserine. *Neuropsychopharmacology, 29(*8), 1573.

Nitsche, M. A., Liebetanz, D., Schlitterlau, A., Henschke, U., Fricke, K., Frommann, K., ... & Tergau, F. (2004). GABAergic modulation of DC stimulation-induced motor cortex excitability shifts in humans. *European Journal of Neuroscience, 19*(10), 2720-2726.

Nitsche, M. A., Seeber, A., Frommann, K., Klein, C. C., Rochford, C., Nitsche, M. S., ... & Paulus, W. (2005). Modulating parameters of excitability during and after transcranial direct current stimulation of the human motor cortex. *The Journal of physiology, 568*(1), 291-303.

Notturno, F., Marzetti, L., Pizzella, V., Uncini, A., & Zappasodi, F. (2014). Local and remote effects of transcranial direct current stimulation on the electrical activity of the motor cortical network. *Human brain mapping, 35*(5), 2220-2232.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass, 4*(10), 906-924.

Ochsner, K., & Kosslyn, S. M. (Eds.). (2013). *The Oxford handbook of cognitive neuroscience, Volume 2: The cutting edges* (Vol. 2). Oxford University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349(*6251), aac4716.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227-252.

Öztekin, I., Curtis, C. E., & McElree, B. (2009). The medial temporal lobe and the left inferior prefrontal cortex jointly support interference resolution in verbal working memory. *Journal of cognitive neuroscience, 21*(10), 1967-1979.

Pahor, A., & Jaušovec, N. (2014). The effects of theta transcranial alternating current stimulation (tACS) on fluid intelligence. *International Journal of Psychophysiology*, *93*(3), 322-331.

Palm, U., Reisinger, E., Keeser, D., Kuo, M. F., Pogarell, O., Leicht, G., ... & Padberg, F. (2013). Evaluation of sham transcranial direct current stimulation for randomized, placebo-controlled clinical trials. *Brain stimulation, 6*(4), 690-695.

Paquette, C., Sidel, M., Radinska, B. A., Soucy, J. P., & Thiel, A. (2011). Bilateral transcranial direct current stimulation modulates activation-induced regional blood flow changes during voluntary movement. *Journal of Cerebral Blood Flow & Metabolism, 31*(10), 2086-2095.

Parazzini, M., Fiocchi, S., Rossi, E., Paglialonga, A., & Ravazzani, P. (2011). Transcranial direct current stimulation: estimation of the electric field and of the current density

in an anatomical human head model. *IEEE Transactions on Biomedical Engineering, 58*(6), 1773-1780.

Parazzini, M., Rossi, E., Ferrucci, R., Liorni, I., Priori, A., & Ravazzani, P. (2014). Modelling the electric field and the current density generated by cerebellar transcranial DC stimulation in humans. *Clinical Neurophysiology, 125*(3), 577-584.

Parazzini, M., Rossi, E., Rossi, L., Priori, A., & Ravazzani, P. (2013). Evaluation of the current density in the brainstem during transcranial direct current stimulation with extra-cephalic reference electrode. *Clinical Neurophysiology, 124*(5), 1039-1040.

Parkin, B.L., Ekhtiari, H., & Walsh, V.F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: A primer. *Neuron, 87*(5), 932-945.

Paulus, W. (2011). Transcranial electrical stimulation (tES–tDCS; tRNS, tACS) methods. *Neuropsychological rehabilitation*, *21*(5), 602-617.

Pelletier, S. J., & Cicchetti, F. (2015). Cellular and molecular mechanisms of action of transcranial direct current stimulation: evidence from in vitro and in vivo models. *International Journal of Neuropsychopharmacology, 18*(2), pyu047.

Penolazzi, B., Pastore, M., & Mondini, S. (2013a). Electrode montage dependent effects of transcranial direct current stimulation on semantic fluency. *Behavioural brain research, 248*, 129-135.

Penolazzi, B., Pastore, M., & Mondini, S. (2013b). Clustering and Switching on Semantic Fluency after Transcranial Direct Current Stimulation (tDCS). *Procedia-Social and Behavioral Sciences, 94*, 39-40.

Perceval, G., Martin, A. K., Copland, D. A., Laine, M., & Meinzer, M. (2017). High-definition tDCS of the temporo-parietal cortex enhances access to newly learned words. *Scientific reports*, *7*(1), 17023.

Pergolizzi, D., & Chua, E. F. (2015). Transcranial direct current stimulation (tDCS) of the parietal cortex leads to increased false recognition. *Neuropsychologia, 66*, 88-98.

Pergolizzi, D., & Chua, E. F. (2016). Transcranial direct current stimulation over the parietal cortex alters bias in item and source memory tasks. *Brain and cognition, 108*, 56-65.

Pergolizzi, D., & Chua, E. F. (2017). Increased contextual cue utilization with tDCS over the prefrontal cortex during a recognition task. *Brain research*, *1655*, 1-9.

Pessoa, L. (2009). How do emotion and motivation direct executive control?. *Trends in cognitive sciences, 13*(4), 160-166.

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J. M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS One, 9*(2), e88674.

Pisoni, A., Mattavelli, G., Papagno, C., Rosanova, M., Casali, A. G., & Romero Lauro, L. J. (2017). Cognitive Enhancement Induced by Anodal tDCS Drives Circuit-Specific Cortical Plasticity. *Cerebral Cortex*, 1-9.

Pisoni, A., Papagno, C., & Cattaneo, Z. (2012). Neural correlates of the semantic interference effect: New evidence from transcranial direct current stimulation. *Neuroscience, 223*, 56-67.

Pisoni, A., Turi, Z., Raithel, A., Ambrus, G. G., Alekseichuk, I., Schacht, A., ... & Antal, A. (2015). Separating recognition processes of declarative memory via anodal tDCS: boosting old item recognition by temporal and new item detection by parietal stimulation. *PloS one, 10*(3), e0123085.

Pisoni, A., Vernice, M., Iasevoli, L., Cattaneo, Z., & Papagno, C. (2015). Guess who? Investigating the proper name processing network by means of tDCS. *Neuropsychologia*, *66*, 267-278.

Polanía, R., Nitsche, M. A., & Paulus, W. (2011). Modulating functional connectivity patterns and topological functional organization of the human brain with transcranial direct current stimulation. *Human brain mapping, 32*(8), 1236-1249.

Polanía, R., Paulus, W., & Nitsche, M. A. (2012). Modulating cortico-striatal and thalamo-cortical functional connectivity with transcranial direct current stimulation. *Human brain mapping, 33(*10), 2499-2508.

Polanía, R., Paulus, W., Antal, A., & Nitsche, M. A. (2011). Introducing graph theory to track for neuroplastic alterations in the resting human brain: a transcranial direct current stimulation study. *Neuroimage, 54*(3), 2287-2296.

Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., & Paulus, W. (2012). The importance of timing in segregated theta phase-coupling for cognitive performance. *Current Biology*, *22*(14), 1314-1318.

Pondé, P. H., De Sena, E. P., Camprodon, J. A., de Araújo, A. N., Neto, M. F., DiBiasi, M., ... & Cosmo, C. (2017). Use of transcranial direct current stimulation for the

treatment of auditory hallucinations of schizophrenia–a systematic review.*Neuropsychiatric disease and treatment*, *13*, 347.

Pope, P. A., & Miall, R. C. (2012). Task-specific facilitation of cognition by cathodal transcranial direct current stimulation of the cerebellum. *Brain stimulation, 5*(2), 84-94.

Prehn, K., & Flöel, A. (2015). Potentials and limits to enhance cognitive functions in healthy and pathological aging by tDCS. *Frontiers in cellular neuroscience, 9.*

Price, A. R., & Hamilton, R. H. (2015). A re-evaluation of the cognitive effects from single-session transcranial direct current stimulation. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8*(3), 663-665.

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.

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *The Journal of Anatomy, 197*(3), 335-359.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*(1), 62-88.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816-847.

Rae, C. D., Lee, V. H. C., Ordidge, R. J., Alonzo, A., & Loo, C. (2013). Anodal transcranial direct current stimulation increases brain intracellular pH and modulates bioenergetics. *International Journal of Neuropsychopharmacology, 16*(8), 1695-1706.

Redfearn, J. W. T., Lippold, O. C. J., & Costain, R. (1964). Preliminary account of the clinical effects of polarizing the brain in certain psychiatric disorders. *The British Journal of Psychiatry, 110*(469), 773-785.

Reinhart, R. M., Cosman, J. D., Fukuda, K., & Woodman, G. F. (2017). Using transcranial direct-current stimulation (tDCS) to understand cognitive processing. *Attention, Perception, & Psychophysics*, *79*(1), 3-23.

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., ... & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences, 106*(5), 1590-1595.

Riggall, K., Forlini, C., Carter, A., Hall, W., Weier, M., Partridge, B., & Meinzer, M. (2015). Researchers' perspectives on scientific and ethical issues with transcranial direct current stimulation: an international survey. *Scientific reports*, *5*, 10618.

Robinson, G., Blair, J., & Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses?. *Brain: a journal of neurology*, *121*(1), 77-89.

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2012). The differing roles of the frontal cortex in fluency tests. *Brain*, *135*(7), 2202-2214.

Romain, P. L. (2015). Conflicts of interest in research: looking out for number one means keeping the primary interest front and center. *Current reviews in musculoskeletal medicine*, *8*(2), 122-127.

Ross, L. A., McCoy, D., Coslett, H. B., Olson, I. R., & Wolk, D. A. (2011). Improved proper name recall in aging after electrical stimulation of the anterior temporal lobes. *Frontiers in aging neuroscience, 3*.

Ross, L. A., McCoy, D., Wolk, D. A., Coslett, H. B., & Olson, I. R. (2010). Improved proper name recall by electrical stimulation of the anterior temporal lobes. *Neuropsychologia, 48*(12), 3671-3674.

Rosso, C., Valabregue, R., Arbizu, C., Ferrieux, S., Vargas, P., Humbert, F., ... & Cohen, L. (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.

Rush, S., & Driscoll, D. A. (1968). Current distribution in the brain from surface electrodes. *Anesthesia & Analgesia, 47*(6), 717-723.

Russo, R., Twyman, P., Cooper, N. R., Fitzgerald, P. B., & Wallace, D. (2017). When you can, scale up: large-scale study shows no effect of tDCS in an ambiguous risk-taking task. *Neuropsychologia, 104*, 133-143.

Sagan, C. (1980). *Broca's brain: Reflections on the romance of science*. Presidio Press.

Saidmanesh, M., Pouretemad, H. R., Amini, A., Nilipor, R., & Ekhtiari, H. (2012). Effects of transcranial direct current stimulation (2mA–20min) in patients with non-fluent aphasia disorder. *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine, 3*(5), 133-144.

Salvador, R., Mekonnen, A., Ruffini, G., & Miranda, P. C. (2010, August). Modeling the electric field induced in a high resolution realistic head model during transcranial

current stimulation. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 2073-2076). IEEE.

Sandars, M., Cloutman, L., & Woollams, A. M. (2015). Taking sides: An integrative review of the impact of laterality and polarity on efficacy of therapeutic transcranial direct current stimulation for anomia in chronic poststroke aphasia. *Neural plasticity,* 2016.

Santarnecchi, E., Brem, A. K., Levenbaum, E., Thompson, T., Kadosh, R. C., & Pascual-Leone, A. (2015). Enhancing cognition using transcranial electrical stimulation. *Current Opinion in Behavioral Sciences*, *4*, 171-178.

Santarnecchi, E., Muller, T., Rossi, S., Sarkar, A., Polizzotto, N. R., Rossi, A., & Kadosh, R. C. (2016). Individual differences and specificity of prefrontal gamma frequency-tACS on fluid intelligence capabilities. *Cortex*, *75*, 33-43.

Santarnecchi, E., Polizzotto, N. R., Godone, M., Giovannelli, F., Feurra, M., Matzen, L., ... & Rossi, S. (2013). Frequency-dependent enhancement of fluid intelligence induced by transcranial oscillatory potentials. *Current Biology*, *23*(15), 1449-1453.

Sarkar, A., Dowker, A., & Kadosh, R. C. (2014). Cognitive enhancement or cognitive cost: trait-specific outcomes of brain stimulation in the case of mathematics anxiety. *Journal of Neuroscience, 34*(50), 16605-16610.

Savine, A. C., Beck, S. M., Edwards, B. G., Chiew, K. S., & Braver, T. S. (2010). Enhancement of cognitive control by approach and avoidance motivational states. *Cognition and Emotion*, *24*(2), 338-356.

Seo, H., & Jun, S. C. (2017). Multi-scale computational models for electrical brain stimulation. *Frontiers in human neuroscience*, *11*, 515.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90.

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, *75*, 27-44.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. Journal of Memory and Language, 54(2), 199-227.

Schooler, J. (2011). Unpublished results hide the decline effect.*Nature*, *470*(7335), 437.

Schutter, D. J., & Wischnewski, M. (2016). A meta-analytic study of exogenous oscillatory electric potentials in neuroenhancement. *Neuropsychologia*, *86*, 110-118.

Scott, R. M., & Wilshire, C. E. (2010). Lexical competition for production in a case of nonfluent aphasia: Converging evidence from four different tasks. Cognitive neuropsychology, 27(6), 505-538.

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.

Shah-Basak, P. P., Norise, C., Garcia, G., Torres, J., Faseyitan, O., & Hamilton, R. H. (2015). Individualized treatment with transcranial direct current stimulation in patients with chronic non-fluent aphasia due to stroke. *Frontiers in human neuroscience, 9*.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in psychology, 5*.

Silvanto, J., Muggleton, N., Lavie, N., & Walsh, V. (2008). The perceptual and functional consequences of parietal top-down modulation on the visual cortex. *Cerebral Cortex, 19*(2), 327-330.

Simons, D. J. (2014). The value of direct replication.*Perspectives on Psychological Science*, *9*(1), 76-80.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory, 6*(2), 174.

Snowball, A., Tachtsidis, I., Popescu, T., Thompson, J., Delazer, M., Zamarian, L., ... & Kadosh, R. C. (2013). Long-term enhancement of brain function and cognition using cognitive training and brain stimulation. *Current Biology*, *23*(11), 987-992.

Snyder, H. R., Feigenson, K., & Thompson-Schill, S. L. (2007). Prefrontal cortical response to conflict during semantic and phonological tasks. *Journal of cognitive neuroscience*, *19*(5), 761-775.

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.

Spielmann, K., Van Der Vliet, R., van de Sandt-Koenderman, W. M. E., Frens, M. A., Ribbers, G. M., Selles, R. W., ... & Holland, P. (2017). Cerebellar Cathodal Transcranial Direct Stimulation and Performance on a Verb Generation Task: A Replication Study. *Neural plasticity*, 2017.

Sprekeler, H. (2017). Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. *Current Opinion in Neurobiology, 43*, 198-203.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *The Neuroscientist, 17*(1), 37-53.

Stagg, C. J., Bachtiar, V., & Johansen-Berg, H. (2011). The role of GABA in human motor learning. *Current Biology, 21*(6), 480-484.

Stagg, C. J., Best, J. G., Stephenson, M. C., O'Shea, J., Wylezinska, M., Kincses, Z. T., ... & Johansen-Berg, H. (2009). Polarity-sensitive modulation of cortical neurotransmitters by transcranial stimulation. *Journal of Neuroscience, 29*(16), 5202-5206.

Stagg, C. J., Lin, R. L., Mezue, M., Segerdahl, A., Kong, Y., Xie, J., & Tracey, I. (2013). Widespread modulation of cerebral perfusion induced during and after transcranial direct current stimulation applied to the left dorsolateral prefrontal cortex. *Journal of Neuroscience, 33*(28), 11425-11431.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59-71.

Swanson, H. L., & McMurran, M. (2017). The impact of working memory training on near and far transfer measures: Is it all about fluid intelligence?. *Child Neuropsychology*, 1-26.

Takeuchi, T., Duszkiewicz, A. J., & Morris, R. G. (2014). The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Phil. Trans. R. Soc. B*, *369*(1633), 20130288.

Talsma, L. J., Kroese, H. A., & Slagter, H. A. (2017). Boosting cognition: Effects of multiple-session transcranial direct current stimulation on working memory. *Journal of cognitive neuroscience*.

Taylor, S. F., Welsh, R. C., Wager, T. D., Phan, K. L., Fitzgerald, K. D., & Gehring, W. J. (2004). A functional neuroimaging study of motivation and executive function. Neuroimage, 21(3), 1045-1054.

Teo, F., Hoy, K. E., Daskalakis, Z. J., & Fitzgerald, P. B. (2011). Investigating the role of current strength in tDCS modulation of working memory performance in healthy controls. *Frontiers in psychiatry, 2*.

Terney, D., Chaieb, L., Moliadze, V., Antal, A., & Paulus, W. (2008). Increasing human brain excitability by transcranial high-frequency random noise stimulation. *Journal of Neuroscience,28*(52), 14147-14155.

Thiel, A., Habedank, B., Herholz, K., Kessler, J., Winhuisen, L., Haupt, W. F., & Heiss, W. D. (2006a). From the left to the right: How the brain compensates progressive loss of language function. *Brain and language, 98*(1), 57-65.

Thiel, A., Schumacher, B., Wienhard, K., Gairing, S., Kracht, L. W., Wagner, R., ... & Heiss, W. D. (2006b). Direct demonstration of transcallosal disinhibition in language networks. *Journal of Cerebral Blood Flow & Metabolism, 26*(9), 1122-1127.

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current opinion in neurobiology, 15(*2), 219-224.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, *23*(3), 513-522.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797.

Thomson, J. M., Doruk, D., Mascio, B., Fregni, F., & Cerruti, C. (2015). Transcranial direct current stimulation modulates efficiency of reading processes. *Frontiers in human neuroscience, 9*.

Thut, G., & Miniussi, C. (2009). New insights into rhythmic brain activity from TMS–EEG studies. *Trends in cognitive sciences*, *13*(4), 182-189.

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.

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of clinical and experimental neuropsychology, 22*(3), 370-378.

Troyer, A. K., & Moscovitch, M. (2006). Cognitive processes of verbal fluency tasks. The quantified process approach to neuropsychological assessment, 143-160.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. N*europsychology, 11*(1), 138.

Tseng, P., Hsu, T. Y., Chang, C. F., Tzeng, O. J., Hung, D. L., Muggleton, N. G., ... & Juan, C. H. (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.

Uhlhaas, P. J., & Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron*, *52*(1), 155-168.

Uhlhaas, P. J., & Singer, W. (2012). Neuronal dynamics and neuropsychiatric disorders: toward a translational paradigm for dysfunctional large-scale networks. *Neuron*, *75*(6), 963-980.

Underwood, E. (2016). Cadaver study challenges brain stimulation methods. *Science, 352*(6284), 397-397.

van Dun, K., Bodranghien, F. C., Mariën, P., & Manto, M. U. (2016). tDCS of the cerebellum: where do we stand in 2016? Technical issues and critical review of the literature. *Frontiers in human neuroscience, 10*.

Vandermeeren, Y., Jamart, J., & Ossemann, M. (2010). Effect of tDCS with an extracephalic reference electrode on cardio-respiratory and autonomic functions. *BMC neuroscience, 11*(1), 38.

Vannorsdall, T. D., Schretlen, D. J., Andrejczuk, M., Ledoux, K., Bosley, L. V., Weaver, J. R., ... & Gordon, B. (2012). Altering automatic verbal processes with transcranial direct current stimulation. *Frontiers in psychiatry*, *3*, 73.

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.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, *2*(4), 229.

Herrmann, C. S., & Demiralp, T. (2005). Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical neurophysiology*, *116*(12), 2719-2733.

Vaseghi, B., Zoghi, M., & Jaberzadeh, S. (2015). A meta-analysis of site-specific effects of cathodal transcranial direct current stimulation on sensory perception and pain. *PloS one, 10*(5), e0123873.

Villamar, M. F., Volz, M. S., Bikson, M., Datta, A., DaSilva, A. F., & Fregni, F. (2013). Technique and considerations in the use of 4x1 ring high-definition transcranial direct current stimulation (HD-tDCS). *Journal of visualized experiments: JoVE*, (77).

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*(1), 103.

Vöröslakos, M., Takeuchi, Y., Brinyiczki, K., Zombori, T., Oliva, A., Fernández-Ruiz, A., ... & Berényi, A. (2018). Direct effects of transcranial electric stimulation on brain circuits in rats and humans. *Nature communications*, *9*(1), 483.

Wagner, T., Fregni, F., Fecteau, S., Grodzinsky, A., Zahn, M., & Pascual-Leone, A. (2007). Transcranial direct current stimulation: a computer-based human model study. *Neuroimage*, *35*(3), 1113-1124.

Walsh, V. Q. (2013). Ethics and social risks in brain stimulation. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 6(*5), 715-717.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*, *90*(3), 1195-1268.

Westwood, S. J., & Romani, C. (2017). Transcranial direct current stimulation (tDCS) modulation of picture naming and word reading: A meta-analysis of single session tDCS applied to healthy participants. *Neuropsychologia*, (104), 234-249

Westwood, S. J., & Romani, C. (2018). Null Effects on Working Memory and Verbal Fluency Tasks When Applying Anodal tDCS to the Inferior Frontal Gyrus of Healthy Participants. *Frontiers in neuroscience*, *12*, 166.

Westwood, S. J., Olson, A., Miall, R. C., & Romani, C. (2017). tDCS modulation of naming in healthy participants: Negative results and still no explanation–a response to a commentary by Gauvin et al.(2017). *Cortex*.

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.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure?. *Applied Neuropsychology: Adult, 23*(1), 29-34.

Wiethoff, S., Hamada, M., & Rothwell, J. C. (2014). Variability in response to transcranial direct current stimulation of the motor cortex. *Brain stimulation, 7*(3), 468-475.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it?. *Frontiers in psychology, 4*.

Wilshire, C. E., & McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology*, *19*(2), 165-186.

Wirth, M., Rahman, R. A., Kuenecke, J., Koenig, T., Horn, H., Sommer, W., & Dierks, T. (2011). Effects of transcranial direct current stimulation (tDCS) on behaviour and electrophysiology of language production. *Neuropsychologia, 49*(14), 3989-3998.

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., ... & Knotkova, H. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. *Clinical Neurophysiology, 127*(2), 1031-1048.

Younger, J. W., Wagner, M. R., & Booth, J. R. (2016). Weighing the Cost and Benefit of Transcranial Direct Current Stimulation on Different Reading Subskills. *Frontiers in neuroscience, 10*.

Yuen, T. G., Agnew, W. F., Bullara, L. A., Jacques, S., & McCreery, D. B. (1981). Histological evaluation of neural damage from electrical stimulation: considerations for the selection of parameters for clinical application. *Neurosurgery, 9*(3), 292-299.

Zaehle, T., Rach, S., & Herrmann, C. S. (2010). Transcranial alternating current stimulation enhances individual alpha activity in human EEG. *PloS one*, *5*(11), e13766.

Zhao, H., Qiao, L., Fan, D., Zhang, S., Turel, O., Li, Y., ... & He, Q. (2017). Modulation of Brain Activity with Noninvasive Transcranial Direct Current Stimulation (tDCS): Clinical Applications and Safety Concerns. *Frontiers in Psychology*, *8*.

Zheng, X., Alsop, D. C., & Schlaug, G. (2011). Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage, 58*(1), 26-33.

Appendices

Appendix 2.1 - Stimuli for Continuous Picture Naming/Word Reading. Permissible synonyms in parenthesis. Words in bold represent those presented in one version of the continuous naming task.

Continuous Picture Naming

Birds: **goose**, **robin, hen (chicken), pigeon, parrot**, rooster (cockerel), duck, pelican, owl, ostrich

Body parts: **tongue, finger, eye, arm, leg**, nail, ear, mouth (lips), foot, nose

Buildings: **shed, barn, lighthouse, church, factory**, cathedral, windmill, skyscraper, tower, castle

Clothing items: **jacket (blazer), socks, sweater (jumper)**, vest, shirt, trousers (chinos), skirt, glove, bathrobe, coat.

Electrical items: **headphones, radio, camera, monitor (screen), printer**, laptop, telephone, speaker, mouse, keyboard.

Farm animals: **horse**, **bull, lamb, calf, donkey (mule)**, sheep, cow, pig, goat, ox.

Flowers: poppy, daffodil, tulip, daisy, dandelion, cactus, sunflower, lavender, rose, lily.

Fruits: **kiwi, apple, lemon, strawberry, pear**, pomegranate, orange, cherries, grapes, melon.

Furniture: **chest, sofa, armchair, stool (chair), bookcase**, chair, cot, chest, wardrobe, table.

Insects: **bee, butterfly, spider, grasshopper (cricket), centipede**, worm, beetle, ant, moth, ladybird.

Kitchen appliances: **blender, whisks, washing machine, oven**, **microwave**, dishwasher, food processor, toaster, kettle, hoover.

Kitchen utensils: **fork**, **colander**, **cup**, **knife**, **frying pan**, spoon, spatula, glass, bowl, pot.

Instruments: **drum**, **guitar**, **flute**, **harp**, **saxophone**, piano, trumpet, violin, clarinet, accordion.

Landscapes: **cliffs**, **river**, **mountain**, **lake**, **sea**, beach, waterfall, iceberg, desert, volcano.

Reptiles: **crocodile**, **toad**, **turtle**, **python**, **iguana**, frog, cobra, lizard, newt, chameleon.

Savoury food: **pizza**, **chicken**, **cracker**, **toast**, **steak**, beans, ham, cheese, bacon, hamburger.

Sea creatures: **crab**, **starfish**, **eel**, **squid**, **lobster**, prawn (shrimp), clam, octopus, oyster, jellyfish.

Stationary: **pen**, **ruler**, **folder**, **paperclip**, **eraser**, pencil, pin, compass, stapler, sharpener.

Sweet food: **ice cream**, **marshmallow**, **brownie**, **cake**, **cookie**, doughnut, croissant, muffin, chocolate cheesecake.

Bathroom items: **soap**, **perfume**, **bud**, **toothbrush**, **toilet paper**, towel, razor, comb, toothpaste, tweezers.

Tools: **axe**, **chisel**, **shears**, **pliers**, **drill**, shovel, mallet, screwdriver, clamp, hammer.

Vegetables: carrot, onion, tomato, lettuce, cauliflower, asparagus, potato, pepper, cucumber, celery.

*Transpor*t: tram, bicycle, plane, tractor, caravan, train, bus, van, helicopter, motorbike.

Safari animals: **hippopotamus**, **camel**, **kangaroo**, **giraffe**, **cheetah**, elephant, tiger, lion, rhino, zebra,

Fillers: nail polish, chain, sword, paintbrush, hourglass, earrings, bag, bauble,

slingshot, match, watering can, binoculars, pillow, brick, coins, dartboard, bow, bottle cap, microscope, mascara brush, cone, spray bottle, clock, suitcase, bat, doll, key, broom, note, brush, ring, chessboard, flyswatter, hose, mousetrap, lighter, bucket, candle, acorn, box, door, peanut, pill, hairband, water bottle.

Appendix 2.2 - Stimulus statistics used for matching parallel versions of the continuous picture naming task; AoA= Age of acquisition from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012), frequency from CELEX Database (Baayen et al., 1995).

List A

	Position										Total	
	1		2		3		4		5		M	SD
	M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD		
Frequency	22	28	17	24	19	38	17	26	19	45	19	33
AoA	6.	3	6	2	6	2	6	2	7	2	6	2
Length	6	2	6	3	6	2	6	2	6.	2	6.	2
Agreement	.7	.9	.9	1.	.8	.7	.6	.9	.7	.7	.7	.9

List B

	Position										Total	
	1		2		3		4		5		M	SD
	M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD		
Frequency	15	20	17	27	18	29	201	38	17	36	18	30
AoA	6	2	6	2	6	2	6	2	6	2	6	2
Length	6	2	6	2	6	3	6	2	7	2	6	2
Agreement	.6	.7	.8	.9	.3	.4	.8	1.	.7	.7	.6	.8

Appendix 2.3 - Stimuli list a Cyclic Blocked Naming.

Cyclic Blocked Naming

Animals: elephant, monkey, panda, rabbit, tiger, zebra.
Bugs: bee, butterfly, fly, grasshopper, mosquito, spider
Body Parts: ear, eye, foot, hand, mouth, nose
Clothing: dress, gloves, hat, sock, tie, trousers
Fruit: apple, banana coconut, grapes, pineapple, melon
Vehicles: ambulance, bicycle, boat, bus, motorbike, train
Appliances: fan, fridges, hoover, phone, tv, washing machine
Birds: duck, owl, peacock, penguin, rooster, swan
People: chef, fireman, nurse, painter, police, teacher,
Tools: hammer, hoe, pliers, saw, scissors, screwdriver
Vegetables: carrot, corn, mushroom, onion, potato, spinach
Food: cake, hamburger, ice cream, pizza, popcorn, turkey

Appendix 2.4 - Stimulus statistics used for matching parallel versions of the Cyclic Blocked Naming task; AoA = Age of acquisition from Kuperman et al., 2012), frequency from CELEX Database (Baayen et al., 1995).

	Categories													
	Animals		*Bugs*		*Body Parts*		*Clothing*		*Fruit*		Vehicles			
List A	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	**Total M**	**Total SD**
Frequency	5.3	1.1	6	1.4	5.1	1.4	5.4	1.7	4.7	0.7	5.7	1.7	**5**	**1.3**
Length	5.8	1.2	6.7	3.3	3.8	0.8	4.8	1.9	7.2	1.9	6.2	2.6	**5.8**	**2.3**
AoA	4.6	0.7	4.5	1.3	3.3	0.4	4.5	1.8	4.8	1.2	5	1.7	**4.5**	**1.3**
Agreement *H*	.1	.2	.9	.7	.3	.1	.3	.4	.2	.2	.4	.3	**.4**	**.4**

	Appliances		*Birds*		*People*		*Tools*		*Vegetables*		*Food*			
List B	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	**Total *M***	**Total *SD***
Frequency	6.1	1.2	5.6	1.5	5.2	1.5	5.3	1.4	4.8	1.4	5.2	1.8	**5.4**	**1.4**
Length	6.2	4.6	5.3	1.9	6	1.3	6.2	3.1	6.2	1.7	6.5	1.9	**6.1**	**2.5**
AoA	5.4	2	5.6	1.1	5.7	1.3	6.2	1.5	5.7	1.9	4.1	0.8	**5.4**	**1.5**
Agreement *H*	.5	.5	.6	.4	.8	.6	.1	.2	.3	.6	.3	.4	**.4**	**.5**

Appendix 2.5 -- Final published version of my response to a commentary made by Gauvin et al. (2017)[14]

2.5.1 Introduction

We would like to thank Drs. Gauvin, Meinzer and de Zubicaray for their commentary on our paper, Westwood, Olson, Miall, Nappo, and Romani (2017). Commentaries are essential to scientific debate because they point out limits to research that may otherwise go unnoticed by the reader. This is especially needed in the field of tDCS, where there is debate regarding its efficacy. Our paper was motivated as an original contribution to this debate, so we gladly accept our chance to respond to their commentary (Gauvin, Meinzer, and de Zubicaray, 2017; hereafter referred to as Gauvin et al.). We first clarify two issues that frame much of what is discussed later.

Firstly, the focus of our investigation was much wider than Gauvin et al. suggested. We wanted to assess whether a single session of anodal tDCS can modify performance on word production tasks in healthy participants, as we made clear throughout, including in the abstract, introduction and above all in the detailed empirical investigation. In our main analyses, we looked at the general effects of anodal tDCS on word reading and picture naming speed and accuracy. Since we failed to find any significant effects in the main analyses, we attempted to find effects with a number of additional analyses of semantic

[14][14]Gauvin, H. S., Meinzer, M., & de Zubicaray, G. I. (2017). tDCS effects on word production: Limited by design? Comment on Westwood et al. (2017). Cortex.

interference effects, of responses at different speeds and by considering possible individual differences in response to tDCS. This amounted to roughly 80 analyses overall, none of which showed significant effects of tDCS. That Gauvin et al. focused on our analyses of semantic interference effects alone misrepresents the aims of our paper.

Secondly, and more importantly, the focus of our paper was not to replicate any specific study. As we explained in the introduction, one aim was to 'try to replicate...findings' that anodal tDCS can modify semantic interference effects, given the inconsistency of these findings. We wanted to give the effects of tDCS the best chance to emerge through different analyses, not to replicate a specific study. There is a difference between a conceptual replication and a direct replication (for discussion, see Cesario, 2014; Schmidt, 2009; Simons, 2014; Stroebe & Strack, 2014). Gauvin et al. failed to appreciate this distinction.

2.5.2 Our response to comments

Gauvin et al. criticized the investigations reported in our paper in terms of the theoretical framework, design, methodology and data analysis. We consider their objections in turn.

2.5.2.1 Issues with theoretical framework

Gauvin et al. said that a key assumption of our study was that the left inferior frontal gyrus (or LIFG) is reliably involved in semantic interference effects. This is not true. In line with the focus of the paper, our key assumption was that the LIFG underpins word production, which is in line with data collected over many years from several lines of research (see Devlin & Watkins, 2007; Indefrey & Levelt, 2004; Lazar & Mohr, 2011; Price, 2000). Exploring the possible modulation of semantic interference effects in picture naming with LIFG stimulation was therefore a necessary aspect of our investigation. In addition, the LIFG has been the focus of a number of previous studies exploring the effects of tDCS on semantic interference, albeit with inconsistent results, as cited in our paper (e.g., Meinzer, Yetim, McMahon, & de Zubicaray, 2016; Pisoni, Papagno, & Cattaneo, 2012; Wirth et al., 2011). Far from being unaware of the current debate regarding the role the LIFG plays in semantic interference, as claimed by Gauvin et al., our hypotheses were clearly formulated in light of this debate. We stated that the hypothesis that interference effects will be reduced with LIFG stimulation depends on the controversial assumption that 'top-down frontal mechanisms

contribute to lexical selection in addition to mechanisms of lateral inhibition intrinsic to the lexical module (see Hamilton & Martin, 2005, 2007 for a discussion)'. See page 66 in paper.

2.5.2.1 Issues with stimulation protocol

Gauvin et al. said that the 'sole experiment involving an attempted replication of prior work' was our Experiment 2, since three previous similar studies had coupled prefrontal tDCS with the cyclic blocked naming task. They then go on to say that while '...Westwood et al. discuss their findings from Experiment 2 in terms of a failure to replicate prior work, it is clear from Table 1 that their tDCS protocol matches none of the previous studies'.

Firstly, as already mentioned, we never set out to directly replicate a specific protocol, but instead we used parameters considered 'best practice'. Thus, our study shared important aspects with other studies without exactly replicating any of them. Across these studies (including ours), all targeted the LIFG (except for: Wirth et al., 2011, which targeted the left dorsolateral prefrontal cortex); all used online stimulation (except for: Pisoni et al., 2012, which used offline stimulation), and all used the same location and size of the reference electrode (except for: Meinzer et al., 2016, which used 100 cm^2 sized reference). One departure of note is that we used a smaller active electrode compared to others (25 vs 35 cm^2), which was motivated by evidence that reducing the size of the active electrode can increase focality (Nitsche et al., 2008), and that this electrode size has been used with success elsewhere (see review by Mancuso, Ilieva, Hamilton, & Farah, 2016). Moreover, because the efficacy of stimulation relies partly on current density (i.e., the current intensity relative to the electrode size), our use of 1.5 mA current meant the current density we applied fell within the range used by the three other studies reported in Table 1 (mA/cm^2 of .03, .04, .06; ours, .06). Thus, we consider a difference in electrode size to be a minor departure from previous protocols, which - if anything - should have increased the likelihood of a significant effect.

Secondly, Gauvin et al. considered the use of online stimulation as an important limitation to our study. Meinzer et al. (2016) interpreted their weak effect of LIFG stimulation as potentially due to the use of online stimulation, and suggested that differences in online/offline stimulation could explain variability in the effects reported with tDCS coupled with blocked cyclic naming. We, like Meinzer et al. (2016), chose online stimulation because

it is thought to target neuronal networks recruited by the task (for a similar argument, see Miniussi, Harris, & Ruzzoli, 2013), because it is considered to produce a stronger increase in excitability compared to offline stimulation (Rae, Lee, Ordidge, Alonzo, & Loo, 2013; Stagg et al., 2013; see also, Martin, Liu, Alonzo, Green, & Loo, 2014) and because positive effects were reported by previous picture naming studies (e.g., Fertonani, Brambilla, Cotelli, & Miniussi, 2014; Ross, McCoy, Wolk, Coslett, & Olson, 2010), including studies listed in Table 1.

Finally, it is certainly true that departures in protocol may result in variation in outcome, as pointed out by Gauvin et al. The problem is that we have not yet identified the conditions in which tDCS can operate reliably, at least within the limit set by our studies (i.e., word production, healthy participants, one stimulation session). Direct replications are a good way to evaluate the reliability/efficacy of protocols, which is why our lab is currently conducting several replications of studies, including Meinzer et al. (2016) and Pisoni et al. (2012). We are continuing in our efforts to establish the conditions under which tDCS is effective.

2.5.3 Issues with design and methodology

Gauvin et al. criticized two main aspects of our methodology, namely the task instructions and the use of both naming and reading tasks.

2.5.3.1 Longer reaction times

Gauvin et al. noticed that our picture naming reaction times (RTs) are longer than in other studies using the continuous picture naming task that they cite (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006, 610-735 msec; Navarrete, Mahon, & Caramazza, 2010; 770-844 msec; Belke, 750-830 msec; our 900-990 msec). They attribute this to our instruction to ask participants to use subordinate names, which, according to them, deviates from previous studies using the continuous picture naming task (e.g., Howard et al., 2006), and may have resulted in a processing cost, as evidence by the fact that our RTs are roughly 150 msec longer than previous studies they cite.

Firstly, we did not use the term 'subordinate names' in the task instructions, but we did ask participants to use precise names, and provided a clear example of what we meant - e.g., correct responses to water-lily could be "water-lily" or "lily" but not "flower" - along with a practice task. These instructions were to prevent participants from applying the same general term to all members of a given category, such as flower, which would have reduced (or abolished) the semantic interference effect. This instruction does not contrast at all with Howard et al. (2006), who designed the original continuous naming task. In fact, it is required by this task. For example, Howard et al. (2006) included pictures of a cap, beret, swordfish, wasp, ladybird, and desk. As with our study, it was important that participants used specific words rather than more generic terms such as hat, fish, insect, or table to name pictures. We simply made this clear to participants.

Secondly, even if we were to grant that there was a processing cost because of our task instructions, would this not be a good thing? It is a rule of human performance that interference effects are normally stronger, not weaker, in more challenging conditions. Consistent with this, previous research has shown that effects of tDCS are more likely when participants are not performing at ceiling (see Berryhill, Peterson, Jones, & Stephens, 2014; Ross et al., 2010). Gauvin et al. failed to mention that we carried out specific analyses to address task difficulty by running separate analyses for responses at different speeds (page 75, Section 3.4). Our assumption was that for harder items - indicated by slower naming speeds - we would find a significant effect of tDCS. We still did not find any effect of tDCS. Thirdly, our longer RTs may reflect the fact that our presentation of the stimuli and trimming procedures allowed longer RTs to be included in our analyses. We displayed pictures for 2500 msec or until a response was made. We excluded RTs shorter than 250 msec and slower than 2.5 standard deviations from the subject mean, as is standard practice. The other studies cited by Gauvin et al. either presented the picture for a shorter time (e.g., 1500 msec in Navarrete et al., 2010 and Belke, 2013) or trimmed longer RTs more (e.g., below 250 and above 2000 msec in Howard et al., 2006). Our longer picture display duration alone would have led to longer RTs. We specifically wanted to include longer RTs in order to carry out more detailed analyses according to speed of responses, as mentioned in the paragraph directly above.

2.5.3.2 Combining reading and naming

Gauvin et al. criticized the fact that we asked participants to perform two tasks - reading and picture naming - which ran sequentially. They argued that there could be possible interactions between reading and naming that cancel out any significant effect of tDCS on picture naming. We find this hard to believe. Firstly, there is no reason to assume that reading should interfere with picture naming, given that the same target words were used in the two tasks. When presented first, reading had the purpose of reducing ambiguity of picture names, in line with common practice. Secondly, and crucially, there was no effect of tDCS on reading in any shape or form. It is not clear how Gauvin et al. imagine the null effect in reading would cancel out an otherwise positive effect in naming.

2.5.3.3 Data analysis

Gauvin et al. suggested that the results we obtained with the continuous naming paradigm were different from previously obtained results. This, supposedly, would put into question the validity of all our experiments, and particularly for Experiment 1c, where we targeted the temporal region, which is implicated in lexico-semantic retrieval, and where stimulation produced significant effects in one of the studies by one of the authors of the commentary (Meinzer et al., 2016). Gauvin et al. pointed out that neither 'lag or session should influence the cumulative interference effect based on previous results (e.g., Belke, 2013)'. Instead, in their reanalysis of data for our control participants - who carried out both sessions without stimulation - Gauvin et al. found an interaction between position, lag and session, which was significant by participants and marginally significant by categories [F_1 (3,78) = 4.07, p = .01, η_p^2 = .14; F_2 (9,250) = 1.88, p = .055, η_p^2 = .06]. They claimed that this interaction makes our results uninterpretable, since 'findings from their Experiment 1b and c with tDCS are confounded by both lag and session'. Gauvin et al. then unpacked this three-way interaction by plotting RTs across positions with respect to lag separately for the pseudo-sham and the pseudo-real session, and query the fact that plots show a quadratic trend as well as a linear trend, which would be a departure from the original findings by Howard et al. (2006).

Three-way interactions are often difficult to interpret, but they do not preclude interpretation. We have carried out more extensive analyses to address the points raised

(for results, see Supplementary Material 1 shown in Appendix 2.6). In six out of the eight analyses, we did not find any three-way interaction of position, lag and session. The only two significant three-way interactions were the same as those found by Gauvin et al. We unpacked them by carrying out separate analyses for each session (pseudo-tDCS and pseudo-sham). For both sessions there was no significant effect of lag and no interaction of lag by position. Instead, an effect of position was highly significant or marginally significant in both sessions [pseudo-tDCS: $F_1(3,72) = 6.61$, $p = .001$, $\eta_p^2 = .22$; $F_2(3,69) = 3.96$, $p = .012$, $\eta_p^2 = .15$; pseudo-Sham: $F_1(3,72) = 2.80$, $p = .046$, $\eta_p^2 = .10$; $F_2(3,69) = 2.29$, $p = .09$, $\eta_p^2 = .09$]. Similarly a linear trend across positions was significant in both sessions [pseudo-tDCS: $F_1(1,24) = 9.67$, $p = .01$, $\eta_p^2 = .29$; $F_2(1,23) = 5.01$, $p = .04$, $\eta_p^2 = .18$; pseudo-sham: $F_1(1,24) = 5.93$, $p = .023$, $\eta_p^2 = .20$; $F_2(1,23) = 5.15$, $p = .033$, $\eta_p^2 = .18$]. We did find a significant quadratic trend by participants and marginally by categories for pseudo-tDCS [$F_1(1,24) = 15.49$, $p = .001$, $\eta_p^2 = .39$; $F_2(1,23) = 5.03$, $p = .04$, $\eta_p^2 = .18$], but not pseudo-sham [$F_1(1,24) = .34$, $p = .56$, $\eta_p^2 = .01$; $F_2(1,23) = .09$, $p = .77$, $\eta_p^2 = .004$]. In Fig. 1 (see Supplementary Material 1), we see that interference diminishes with longer lags, particularly at lag 8. This finding is not unique to our data, and was noted recently by Schnur (2014), who reported a reduced interference effect with lags of 8-50.

Thus, overall, our results are strongly consistent with the original results by Howard et al. (2006). Three-way interactions are often difficult to interpret especially when they are not in a predicted and/or theoretically meaningful direction. The only two three-way interactions we found are likely to be an uninteresting result which could have happened by chance. There is no indication that the accumulation of interference is systematically influenced by lag and/or session. Gauvin et al. offer no explanation for the three-way interactions and no explanation of how they could have eliminated any significant effect of tDCS, especially since they occurred in a control group that did not receive tDCS.

2.5.4 Other issues with Gauvin et al.

In their conclusion, Gauvin et al. said that we 'interpret [our] data...as an unsuccessful replication and as evidence that the tDCS technique lacks overall efficiency', and that this has 'broader implications for the field. For instance grant reviewers, who are often not expert

in the specific field of an application, might be unduly influenced by assertions of 'failed replications' and dismiss the importance of continuing the proposed research'.

Gauvin et al.'s conclusion showed a puzzling misinterpretation of our results. We do not interpret our findings as either a direct replication, or as evidence that tDCS '...lacks overall efficiency'. We describe our work as failing to find positive effects of tDCS in certain conditions, which we are very careful to specify, and we also outline conditions where tDCS is and/or could be potentially effective, with recommendations for future research. An honest assessment of the tDCS literature shows that cognitive effects of tDCS are generally unreliable or weak, especially with healthy participants in single applications, an opinion shared by many researchers (see opinion survey by Riggall et al., 2015). We firmly stand behind our claim that studies have failed to show that tDCS is consistently able to modulate cognition in healthy participants.

Gauvin et al. listed valuable strategies to increase the rigour of the tDCS field, such as direct replication and pre-registration. An important additional strategy, however, is carrying out meta-analyses which collate disparate findings and increase power. We have recently carried out such a meta-analysis to assess the foundational claim that tDCS can modify picture naming and word reading (Westwood & Romani, 2017). We reviewed 14 papers measuring tDCS effects across a total of 96 conditions. Our intentions were to a) quantify effects of conventional protocols that target language regions (e.g., left hemisphere anodal tDCS administered to temporal/frontal areas), either under normal conditions or conditions that induce semantic interference; b) identify parameters which may mod- erate the size of the tDCS effect (within conventional stimulation protocols), such as stimulation timing, current density and duration, and atypical protocols (e.g., right hemisphere anodal tDCS or left/right hemisphere cathodal tDCS). In all analyses there was no significant effect of tDCS on overall naming accuracy or speed and no influence on interference effects (these analyses included the studies mentioned in Table 1 presented in Gauvin et al.). No overall effect of tDCS was found whether or not our studies from Westwood et al. (2017) were included.

Negative results do not mean that research on tDCS should be abandoned, but that efforts should be placed in finding conditions where tDCS is indeed effective. We find it ironic that

Gauvin et al. took issue with the justifiably sceptical tone of our paper because it might 'prevent the field from progressing as funding is diverted elsewhere, and contribute to the perception of experimental psychology as experiencing a replication "crisis"'. Surely unduly inflating the efficacy of tDCS will have an even worse outcome, since time, energy and money will be wasted, and attention diverted from investigating those conditions in which tDCS may in fact be reliable and effective. Such negative repercussions will no doubt damage the reputation of tDCS research (including experimental psychology), and raise important moral and ethical questions, as eloquently delineated by Vincent Walsh, a prominent researcher in the field of non-invasive brain stimulation (Walsh, 2013). Before we conclude, we would like to end our response with a few choice words from Walsh (2013):

'When my friends and colleagues say that "tDCS is a non- invasive brain stimulation (NIBS) neuromodulatory technique, whose clinical applications to treat pathological neuropsychiatric conditions are rapidly growing [Santarnecchi, Feurra, Galli, Rossi, & Rossi, 2013]." I think they fall into a language trap (in which we all find ourselves) of confusing claims with reality. ... I am all for hope, but when it crosses the line into faith, it becomes an unthinking vehicle. ... [One] consequence of the hype is that the noise may mask important findings. We saw the effects of this with depression and TMS, the advance of which was slowed by premature claims and masked by claims about the utility of TMS in just about every neurological and psychiatric condition. ... We would do better to simply be more honest about the limits of our findings'.

2.5.5 Conclusion

We would again like to thank Gauvin et al. for commenting on our work, although we take issue with the fact that they repeatedly misrepresented our work. In our response, we have made clear that their criticisms are without merit and they fail to offer adequate alternative explanations for the null effects we report in Westwood et al. (2017). Gauvin et al. (wrongly) characterized our study as a direct replication and then criticized us for carrying out original experiments rather than trying to exactly replicate previous studies. We see carrying out a fresh series of experiments to assess the ability of tDCS to modulate word production as an important contribution. We find no value in the methodological criticisms raised by Gauvin et al., since our paradigms followed very closely those previously reported in the literature

and we obtained very similar behavioural results. This makes us very confident that our paradigms were sensitive to the effects of semantic interference, which we intended to modulate with tDCS.

Finally, we agree that we provided less evidence regarding stimulation of the temporal lobe and more evidence would be desirable. We also agree that if tDCS research is to rise to the rigorous standards that is demanded if potential benefits are to be harvested, then direct replication as well as conceptual replication studies are key. As we said in our conclusion to our paper, one should no longer assume 'a level of reliability that is not there' but rather take the 'unreliability of tDCS results... as a starting point and as a challenge that needs addressing'.

Our lab is already conducting a direct replication study to assess the effectiveness of tDCS on fluency tasks. Following this commentary, we will also carry out a replication of Meinzer et al. (2016) and Pisoni et al. (2012). These two studies have targeted the left temporal regions, yet both find discrepant results. Clearly differences in protocol may have contributed to differences in outcome, or it may be that tDCS is not reliable. A replication will not only contribute to the exchange above, but also to the debate about whether tDCS can in fact modulate word production and, especially, semantic interference effects.

Appendix 2.6 - Supplementary Material for my response to a commentary made by Gauvin et al. (2017)

2.6.1 Supplementary Material

The significant three-way interaction reported by Gauvin et al. prompted a re-analysis of all our data with lag included as a factor. We analysed naming reaction times both by participants (F_1) and by categories (F_2), with *Position* (2 to 5), *Lag* (2 to 8), and *Session* (pseudo-tDCS versus pseudo-Sham, for control participants; tDCS versus Sham, for experimental participants), as within participants factors. These analyses necessarily exclude position 1, because there could be no effect of lag for this position (see Table 2 and Figure 1 for results). Importantly, we were unable to carry out analyses by categories for Experiment 1A (control and experimental data). In their original study, Howard, Nickels, Coltheart, and Cole-Virtue (2006) generated 24 lists of stimuli, each with a different sequence of lags for the 24 different semantic categories, with a different set of items in the five ordinal positions. Unfortunately, as we state in our paper, for Experiment 1A, we created two parallel versions of the continuous naming task, with only one list of stimuli for each version, which was given to all participants. This meant that, unlike Howard et al. (2006), our by categories analysis had data missing from all categories for certain combinations of lag and position. In Experiment 1B and 1C, however, we decided to generate 24 different lists, like Howard et al. (2006). This is why we were able to run a by categories analyses.

The effect of position was significant in all 8 analyses, but one (by participants in Experiment 1C; here the effect was significant only with the inclusion of the first position). A linear trend was significant in all conditions. The effect of lag was significant only in 2 of the 8 analyses; by participants for Experiment 1A for experimental and control participants, which also showed a significant *Lag* x *Position* interaction. These effects are likely to be due to random variation. There were no other significant *Lag* x *Position* interactions. As the plots in Figure 1 show, consistent with previous results, there is no systematic influence of lag.

In 6 out of the 8 analyses, we did not find any three-way *Position x Lag x Session* interaction. The only two significant three-way interactions were those found by Gauvin et al. – ($F1(9,216) = 2.44$, $p = .01$, $\eta p^2 = .09$; $F2(9,207) = 1.99$, $p = .04$, $\eta p^2 = .08$)[15]. We unpacked them by carrying out separate analyses for each session (pseudo-tDCS and pseudo-sham,). For both sessions there was no significant effect of *Lag* and no interaction of *Lag* x *Position*. Instead, an effect of *Position* was highly significant or marginally significant in both sessions (pseudo-tDCS: $F_1(3, 72) = 6.61$, $p = .001$, $\eta p^2 = .22$; $F_2 (3,69) = 3.96$, $p = .012$, $\eta p^2 = .15$; pseudo-Sham: $F_1(3,72) = 2.80$, $p = .046$, $\eta p^2 = .10$; $F_2 (3,69) = 2.29$, $p = .09$, $\eta p^2 = .09$). Similarly a linear trend across positions was significant in both sessions (pseudo-tDCS: $F_1(1,24) = 9.67$, $p = .01$, $\eta p^2 = .29$; $F_2(1,23) = 5.01$, $p = .04$, $\eta p^2 = .18$; pseudo-sham: $F_1(1,24) = 5.93$, $p = .023$, $\eta p^2 = .20$; $F_2(1,23) = 5.15$, $p = .033$, $\eta p^2 = .18$). We do find a significant quadratic trend by participants and marginally by categories for pseudo-tDCS ($F_1(1,24) = 15.49$, $p = .001$, $\eta p^2 = .39$; $F_2(1,23) = 5.03$, $p = .04$, $\eta p^2 = .18$), but not pseudo-sham ($F_1(1,24) = .34$, $p = .56$, $\eta p^2 = .01$; $F_2(1,23) = .09$, $p = .77$, $\eta p^2 = .004$).

[15]You may note that our values are slightly different than Gauvin et al's. Firstly, there was a difference in the reported degrees of freedom. We do not recognise the degrees of freedom the authors report, and we assume this is down to a clerical error on their part. Second, there is a difference in the reported values for *F* and *p*. This might be due to a rounding error, or to differences in software. We use Statistical Package for the Social Sciences (or SPSS; IBM SPSS Statistics, Version 21), but the software used by Gauvin et al was not reported. This matters because the data matrix for the control data for Experiment 1B and C had cases with missing values (1 in by participants; 5 in by categories), presumably incorrect, too slow or too fast reaction times. SPSS conducts a list- and pair-wise deletion for cases with missing values (Rubin, Witkiewitz, Andre & Reilly, 2007). We replaced missing values with an average for a given lag, position and session combination.

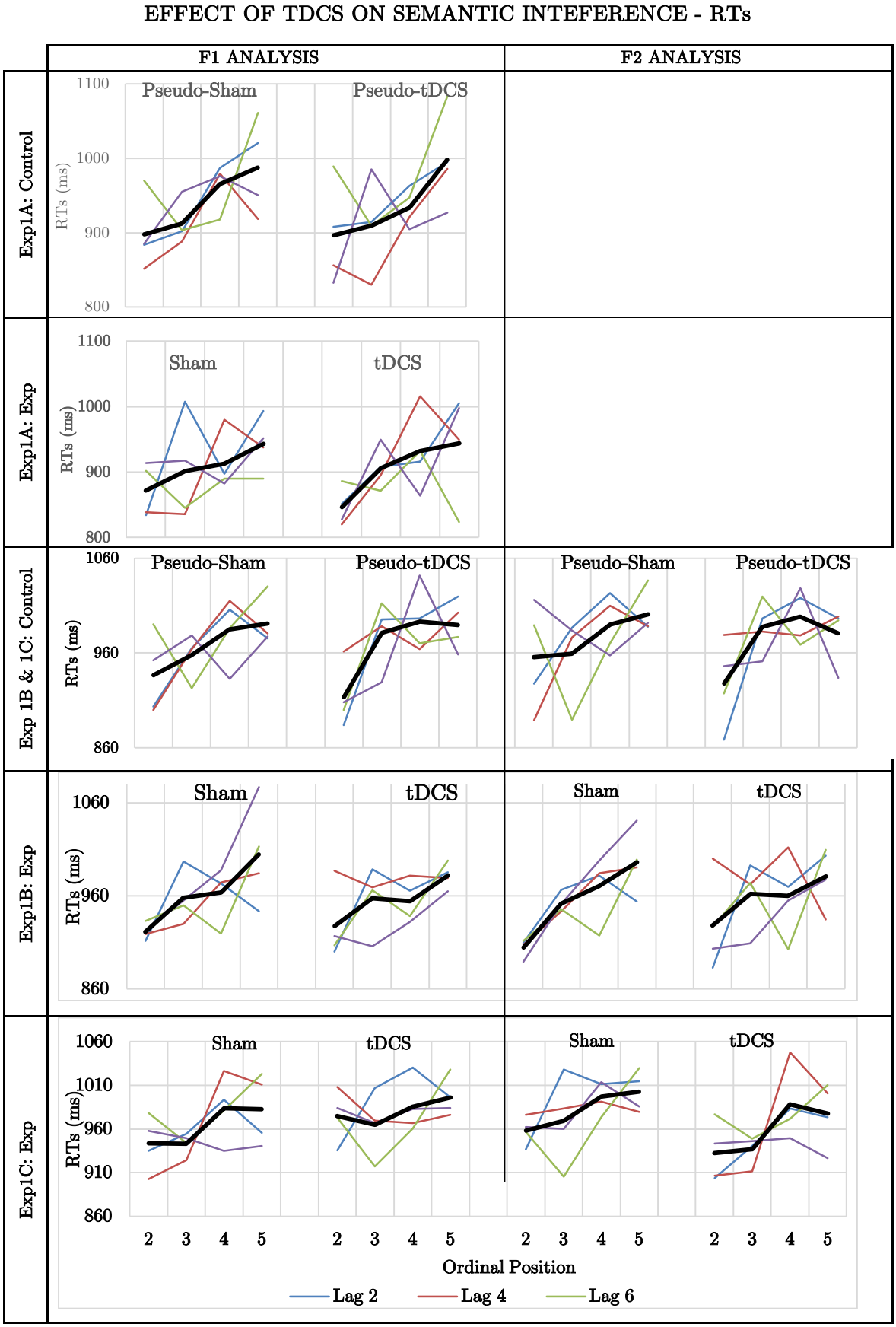


Figure 1. Plots for naming latencies by participants (F1 analysis, left) and by categories (F2 analysis, right) across all experiments, participant groups and stimulation conditions.

Table 2. Statistics for all main effects and interactions for naming latencies by participants (F1) and by categories (F2) across all experiments and participant groups. Significant results highlighted in grey.

		tDCS	Lag	Pos	tDCS*Lag	tDCS*Pos	Lag*Pos	tDCS*Lag*Pos	Linear Trend for Main Effect of Pos	Quadratic Trend for Main Effect of Pos
Cont 1A	F1	$F^{(1,17)}$=.05, *p=.83,* ηp^2*=.002*	$F^{(3,81)}$=3.69, *p=.02,* ηp^2*=.12*	$F^{(3,81)}$=19.65, *p<.001,* ηp^2*=.42*	$F^{(3,81)}$=1.57, *p=.20,* ηp^2*= .06*	$F^{(3,81)}$=.90, *p=.45,* ηp^2*=.03*	$F^{(9,243)}$=6.98, *p<.001,* ηp^2*=.21*	$F^{(9,243)}$=1.09, *p=.37,* ηp^2*=.04*	$F^{(1,27)}$=58.45, *p<.001,* ηp^2*=.68*	$F^{(1,27)}$=1.06, *p=.31,* ηp^2*=.04*
Exp 1A	F1	$F^{(1,17)}$=.00, *p=.995,* ηp^2*<.001*	$F^{(3,51)}$=5.74, *p=.02,* ηp^2*=.25*	$F^{(3,51)}$=14.69, *p<.001,* ηp^2*=.46*	$F^{(3,51)}$=.55, *p=.65,* ηp^2*=.46*	$F^{(3,51)}$=.88, *p=.46,* ηp^2*=.05*	$F^{(9,153)}$=5.96, *p<.001,* ηp^2*=.26*	$F^{(9,153)}$=1.50, *p=.15,* ηp^2*=.08*	$F^{(1,17)}$=36.12, *p<.001,* ηp^2*=.68*	$F^{(1,17)}$=1.59, *p=.23,* ηp^2*=.09*
Cont 1B/C	F1	$F^{(1,24)}$=.01, *p=.94,* ηp^2*<.001*	$F^{(3,72)}$=.33, *p=.80,* ηp^2*=.01*	$F^{(3,72)}$=8.84, *p<.001,* ηp^2*=.27*	$F^{(3,72)}$=.45, *p=.72,* ηp^2*=.02*	$F^{(3,72)}$=.83, *p=.48,* ηp^2*=.03*	$F^{(9,216)}$=.55, *p=.84,* ηp^2*=.02*	$F^{(9,216)}$=2.44, *p=.01,* ηp^2*=.09*	$F^{(1,24)}$=17.43, *p<.001,* ηp^2*=.42*	$F^{(1,24)}$=6.36, *P=.02,* ηp^2*=.21*
	F2	$F^{(1,23)}$=.05, *p=.82,* ηp^2*=.00*	$F^{(3,69)}$=.01, *p=1.0,* ηp^2*=.001*	$F^{(3,69)}$=4.09, *p=.01,* ηp^2*=.15*	$F^{(3,72)}$=.65, *p=.58,* ηp^2*=.03*	$F^{(3,69)}$=1.78, *p=.16,* ηp^2*=.07*	$F^{(9,207)}$=1.43, *p=.18,* ηp^2*=.06*	$F^{(9,207)}$=1.99, *p=.04,* ηp^2*=.08*	$F^{(1,23)}$=7.09, *p=.01,* ηp^2*=.24*	$F^{(1,23)}$=2.35, *p=.14,* ηp^2*=.09*
Exp 1B	F1	$F^{(1,19)}$=.13, *p=.72,* ηp^2*=.01*	$F^{(3,57)}$=.32, *p=.81,* ηp^2*=.021*	$F^{(3,57)}$=6.937, *p<.001,* ηp^2*=.27*	$F^{(3,57)}$=2.40, *p=.08,* ηp^2*=.11*	$F^{(3,57)}$=.37, *p=.77,* ηp^2*=.02*	$F^{(9,171)}$=1.41, *p=.19,* ηp^2*=.07*	$F^{(9,171)}$=.76, *p=.66,* ηp^2*=.04*	$F^{(1,19)}$=22.48, *P<.001,* ηp^2*=.54*	$F^{(1,19)}$=.003, *p=.96,* ηp^2*<001*
	F2	$F^{(1,23)}$=.002, *p=.96,* ηp^2*<.001*	$F^{(3,69)}$=.14, *p=.94,* ηp^2*=.01*	$F^{(3,69)}$=3.43, *p=.02,* ηp^2*=.13*	$F^{(3,69)}$=1.82, *p=.15,* ηp^2*=.07*	$F^{(3,69)}$=.81, *p=.49,* ηp^2*=.03*	$F^{(9,207)}$=1.27, *p=.26,* ηp^2*=.05*	$F^{(9,207)}$=.80, *p=.62,* ηp^2*=.03*	$F^{(1,23)}$=6.59, *p=.02,* ηp^2*=.22*	$F^{(1,23)}$=.17, *p=.68,* ηp^2*=.01*
Exp 1C	F1	$F^{(1,17)}$=.77, *p=.39,* ηp^2*=.04*	$F^{(3,51)}$=.39, *p=.76,* ηp^2*=02*	$F^{(3,51)}$=2.51, *p=.07,* ηp^2*=.13*	$F^{(3,51)}$=1.52, *p=.22,* ηp^2*=.08*	$F^{(3,51)}$=.35, *p=.79,* ηp^2*=.02*	$F^{(3,51)}$=1.18, *p=.31,* ηp^2*=.07*	$F^{(3,153)}$=.78, *p=.64,* ηp^2*=.04*	$F^{(1,17)}$=4.83, *p=.04,* ηp^2*=.22*	$F^{(1,17)}$=.23, *p=.64,* ηp^2*=.01*
	F2	$F^{(1,23)}$=1.36, *p=.26,* ηp^2*=.06*	$F^{(3,69)}$=1.46, *p=.23,* ηp^2*=.06*	$F^{(3,69)}$=3.66, *p=.02,* ηp^2*=.14*	$F^{(3,69)}$=1.71, *p=.17,* ηp^2*=.07*	$F^{(3,69)}$=.65, *p=.59,* ηp^2*=.03*	$F^{(9,207)}$=1.78, *p=.07,* ηp^2*=.07*	$F^{(9,207)}$=1.28, *p=.25,* ηp^2*=.05*	$F^{(1,23)}$=7.29, *p=.01,* ηp^2*=.24*	$F^{(1,23)}$=.37, *p=.55,* ηp^2*=.02*

Appendix 3.1 - Sample of studies included in the review, with details on stimulation parameters and a summary of findings

Author	*n*	Design	A,C	Timing	mA	mA/cm^2	Anode/cm^2	Mins	Target	Ref	RTs/Acc
Boehringer et al. (2013)	39	W	C	Off	2	0.08	25	25	RCereb	RC	RTs
Summary: the right cerebellum was targeted with cathodal tDCS, and participants were tested on various tasks before and after stimulation. One task involved reading aloud a list of 42 colour words as fast as possible. Comparisons between pre and post scores showed no significant effect on reading speeds.											
Fertonani et al. (2010), Exp 1	12	W	A,C	Off	2	0.057	35	8	LdlPFC	RS	RTs
Fertonani et al. (2010), Exp 2	12	W	A,C	Off	2	0.057	35	10	LdlPFC	RS	RTs
Summary: in two separate experiments, the authors measured picture naming accuracy and reaction times for two sets of stimuli (actions and objects) following anodal or cathodal tDCS to the left dlPFC. *Experiment 1:* naming accuracy and reaction times did not show any significant effect of tDCS. *Experiment 2:* participants were faster after anodal tDCS but not cathodal tDCS, with no effect on accuracy scores.											
Fertonani et al. (2014)	20	W	A	Off/On	2	0.057	35	10	LdlPFC	RS	RTs
Summary: old and young participants were given anodal tDCS online and offline in separate conditions. Picture naming reaction times, but not accuracy, were significantly faster both for online and offline tDCS in younger adults and only for online tDCS in older adults.											
Henseler et al. (2014)	36	W	A	On	2	0.08	25	25	LIFG, LMTG	CO	RTs, Acc
Summary: anodal tDCS was applied to the left inferior frontal gyrus (LIFG) or middle temporal gyrus (MTG) whilst participants performed the picture-word interference task. Word distractors were presented at different picture stimuli onset asynchronies (SOAs) to measure the differential effects of related distractors (i.e., for interference, 100ms SOAs; for facilitation, 300ms SOAs). There was no effect of tDCS in any conditions in terms of percentage of errors and reaction times.											
Jeon and Han (2012)	8	B	A	Off	1	0.029	35	20	L/RdlPFC	CO	RTs, Acc
Summary: participants were administered anodal tDCS to the right or left dlPFC. Subjects performed a series of tasks, including the *Stroop task* and the *Korean-Boston Naming Test (KBNT)*. The Stroop task required participants to name colour names printed in black (*word condition*), Xs printed in colours (*colour condition*), and colour words in incongruent colour ink (*interference condition*). The KBNT involved parallel versions with 60 items divided by 4 test periods). Participants were asked to name pictures, and hints were given whenever necessary. Performance was measured by the number of hints provided (e.g., 4-points with no hints; 3-points with meaningful hints, 1-point with first syllable hint; ½ a point for second syllable hint; 0-points for no response; 60 is the maximum score overall). For the Stroop task, reaction times in the *word condition* were reduced for left anodal tDCS compared to pre-stimulation. Left and right anodal stimulation lead to a significantly diminished interference effect compared to pre-stimulation values both immediately after stimulation and two-weeks later. For the *KBNT*, the left anodal tDCS group improved from pre-stimulation immediately and two-week after stimulation.											
Meinzer et al. (2016)	24	B	A	On	1	0.029	35	20	LSTG, LIFG	CO	RTs
Summary: anodal tDCS was applied to the LIFG and posterior middle and superior temporal gyri (pMTG/STG) whilst participants performed the cyclic blocked naming task. The results showed no overall effects of anodal tDCS relative to sham. However, LIFG stimulation reduced the magnitude of semantic interference in early cycles, whilst pMTG/STG reduced interference from the second cycle onwards.											
Pisoni et al. (2012), Exp 1	12	W	A	Off	2	0.057	35	20	LSTG	CO	RTs
Pisoni et al. (2012), Exp 2	12	W	A	Off	2	0.057	35	20	LIFG	CO	RTs
Summary: participants performed the cyclic blocked naming task immediately after anodal tDCS to the LIFG or left Wernicke's area (LMTG) in two separate experiments. *Experiment 1 (LMTG tDCS)*: participants were much slower overall, and responses were more slower for semantically related picture sets compared to sham – i.e., anodal tDCS boosted semantic interference. *Experiment 2 (LIFG tDCS):* though there was no overall facilitatory effect of anodal tDCS there was a significant reduction in the size of semantic interference effect. There was no statistical difference between related and unrelated sets for real stimulation, but there was in sham stimulation.											
Pope and Miall (2012)	22	B	A,C	Off	2	0.08	25	20	RCereb	LD	RTs
Summary: after anodal or cathodal tDCS to the right cerebellum, participants repeatedly performed noun and verb reading and verb generation tasks. Each task was performed 6 times (5 times with the same stimuli then on the 6th time new stimuli were presented). Only verb generation showed a significant effect, with enhancement in terms of learning rate (i.e., change between the first and last repetition for post-tDCS scores) for cathodal tDCS compared to sham (and anodal tDCS).											
Ross et al. (2010)	15	W	A	On	1.5	0.043	35	15	RATL	CC	Acc
Summary: anodal tDCS was applied to the right or left anterior temporal lobes (ATLs) whilst naming pictures of faces and places. Though there was no overall reduction in reaction times or improvement in accuracy, difficult to name items (responses > 5seconds) showed a benefit of anodal tDCS to the right ATL, with a 11% increase when naming pictures of faces.											
Sparing et al. (2008)	15	W	A,C	Off/On	2	0.057	35	7	CP5/6	CO	RTs

Summary: in separate stimulation sessions, anodal tDCS was applied to left or right Wernicke's area or cathodal tDCS was applied to the left Wernicke's area. Picture naming speeds and accuracy were measured during stimulation, immediately, 5 and 10mins after stimulation. There was a significant reduction in naming immediately following left Wernicke anodal tDCS compared to sham. Accuracy was unaffected by tDCS, but data was not reported.

Turkletaub et al. (2012)	25	W	A	Off	1.5	0.06	25	20	LpTC	An	Acc

Summary: anodal tDCS was administered to the posterior temporal cortex before participants performed several assessments of word reading abilities (e.g., Woodcock Reading Mastery Test-Revised-Normative Update or WRMT; Test of Word Reading Efficiency or TOWRE). Anodal tDCS improved sight reading efficiency (i.e., reading list of words as fast as possible) particularly in below average readers.

Westwood et al. (2017), Exp 1A	18	W	A	On	1	0.11	9	15	LIFG	CO	RTs, Acc
Westwood et al. (2017), Exp 1B	20	W	A	On	1.5	0.06	25	25	LIFG	CO	RTs, Acc
Westwood et al. (2017), Exp 1C	18	W	A	On	1.5	0.06	25	25	LpMTG	CO	RTs, Acc
Westwood et al. (2017), Exp 2	20	W	A	On	1.5	0.06	25	25	LIFG	CO	RTs

Summary: across four experiments participants were administered anodal tDCS whilst performing picture naming (continuous picture naming and cyclic blocked naming) and word reading tasks. The study failed to find any effect of tDCS in all experiments either in terms of RTs or percentage errors.

Wirth et al. (2011)	20	W	A	Off	1.5	0.043	35	37	LdlPFC	RS	RTs

Summary: anodal tDCS targeted the left dlPFC whilst participants performed the cyclic blocked naming task. A simple picture naming task was then performed post-stimulation. There was no overall effect of tDCS, but there was a significant reduction in the semantic interference in terms of RTs. Simple picture naming performance remained unchanged.

Younger et al. (2016)	32	B	A	Off	1.5	0.06	25	20	L/RIPL	CO	Acc

Summary: participants were administered either anodal tDCS or sham to the left or right inferior parietal lobe (IPL) after which they were asked to perform two measures of reading ability: single word reading efficiency and rhyme judgment. Participants who received left IPL tDCS improved in terms of reading efficiency relative to sham, but improved less on rhyme judgment relative to right IPL tDCS.

Legend: A = anodal; Acc = percentage errors; An = analogue; ATL = anterior temporal lobe; B = between-participants; C = cathodal; CC = contralateral cheek; CO = contralateral supraorbital area; CP5 = Wernicke's area; CP6 = Wernicke's area analogue; dlPFC = left dorsolateral prefrontal cortex; IPL = inferior parietal lobe; L = left; LD = left deltoid; IFG = inferior frontal gyrus; MTG = middle temporal gyrus; n = number of participants; On = online; Off = offline; pTC = posterior temporal cortex; pMTG = posterior middle temporal gyrus; R = right; RC = right cheek; RCereb = right cerebellum; Ref = reference; RS = right shoulder; RTs = reaction times; S = sham; STG = superior temporal gyrus; W = within-participants

Appendix 3.2 – Listing of studies included in our meta-analysis with between-participants effect size estimates (as often reported in the literature) and our own composite effect size estimate based on a within-design assumption where appropriate. Note, for Moderator analyses, different effects sizes were used for online and offline stimulation since this was a parameter of interest. Effects were aggregated for the Primary and Secondary analyses to avoid violations of the independence assumption. We indicate significant effects as reported by authors in the paper (with Y or N) and as we report based on our composite effect size estimates (underlined)

Table 1 – Studies used in Primary and Moderator analyses with reaction times as the dependent variable

								Between-participants estimate			Composite effects size estimate				
Author	**Condition**	***n***	M_{Sham}	SD_{Sham}	M_{tDCS}	SD_{Sham}	**Sig.**	***g***	CI_{Lower}	CI_{Upper}	M_{Sham} - M_{tDCS}	***V***	***g***	CI_{Lower}	CI_{Upper}
Fertonani et al. (2010), Exp 1	Off, LdlPFC, Objects	12	739	81	731	99	N	0.08	-0.66	0.83	22	463	0.28	-0.25	0.82
	Off, LdlPFC, Actions	12	907	104	871	78	N	0.36	-0.41	1.14					
Fertonani et al. (2010), Exp 2	Off, LdlPFC, Objects	12	617	51	590	47	Y	0.51	-0.29	1.31	38	262	<u>0.62</u>	<u>0.04</u>	<u>1.2</u>
	Off, LdlPFC, Actions	12	789	100	741	58	Y	0.55	-0.26	1.35					
Fertonani et al. (2014)	Off, LdlPFC, Objects	20	757	72	710	72	Y	0.63	-0.03	1.28	28	133	<u>0.52</u>	<u>0.07</u>	<u>0.97</u>
	Off, LdlPFC, Actions	20	585	57	576	56		0.15	-0.45	0.75					
	On, LdlPFC, Objects	20	757	72	720	69	Y	0.5	-0.13	1.14	22	129	0.42	-0.02	0.86
	On, LdlPFC, Actions	20	585	57	578	55		0.12	-0.48	0.72					
Henseler et al. (2014)	On, LIFG, Ass	36	683	66	692	66	N	-0.13	-0.59	0.32	-4	62	-0.08	-0.4	0.24
	On, LIFG, AssUn	36	706	66	719	72	N	-0.18	-0.64	0.27					
	On, LIFG, Sem	36	763	72	764	72	N	-0.01	-0.47	0.44					
	On, LIFG, SemUn	36	730	60	734	60	N	-0.07	-0.52	0.39					
	On, LMTG, Ass	36	683	66	695	72	N	-0.17	-0.77	0.43					
	On, LMTG, AssUn	36	706	66	704	48	N	0.03	-0.56	0.63					
	On, LMTG, Sem	36	763	72	762	72	N	0.01	-0.58	0.61					
	On, LMTG, SemUn	36	730	60	725	48	N	0.09	-0.51	0.68					
Jeon and Han (2012)	Off, LdlPFC, Stro	8	11	3	10	2	N	0.37	-0.56	1.31	1	2	0.37	-0.56	1.31

Meinzer et al. (2016)	On, LIFG, Related	24	643	88	643	87	N	0	-0.55	0.55	-5	206	0.03	-0.36	0.42
	On, LIFG, Mixed	24	606	80	616	94	N	-0.11	-0.66	0.44					
	On, LSTG, Related	24	643	88	625	93	N	0.19	-0.36	0.74	8	198			
	On, LSTG, Mixed	24	606	80	608	83	N	-0.02	-0.57	0.52					
Pisoni et al. (2012)	Off, LSTG, Related	12	604	59	642	66	**Y**	-0.56	-1.37	0.25	-26	198	-0.5	-1.06	0.07
	Off, LSTG, Mixed	12	591	55	605	62	N	-0.22	-0.98	0.53					
	Off, LIFG, Related	12	646	48	621	59	N	0.43	-0.35	1.22	16	173	0.33	-0.22	0.87
	Off, LIFG, Mixed	12	618	45	612	66	N	0.1	-0.65	0.85					
Sparing et al. (2008)	On, CP5	15	531	367	525	306	N	0.09	-0.59	0.77	6	205	0.1	-0.38	0.58
	Off, CP5	15	528	412	499	325	**Y**	0.4	-0.3	1.11	29	247	0.45	-0.05	0.96
	Off, CP5, 5mins	15	535	390	523	287	N	0.18	-0.5	0.87	12	208	0.2	-0.28	0.69
	Off, CP5, 10mins	15	524	435	529	268	N	-0.07	-0.75	0.6	-5	240	-0.08	-0.56	0.4
Westwood et al. (2017), 1A	On, LIFG, Naming	18	896	144	894	126	N	0.01	-0.61	0.64	2	824	0.02	-0.43	0.46
	On, LIFG, Reading	18	497	82	486	71	N	0.14	-0.49	0.76	11	265			
Westwood et al. (2017), 1B	On, LIFG, Naming	20	946	118	945	106	N	0.01	-0.59	0.6	1	508	0.01	-0.41	0.43
	On, LIFG, Reading	20	541	103	537	66	N	0.04	-0.55	0.64	4	340			
Westwood et al. (2017), 1C	On, LpMTG, Naming	18	955	111	969	107	N	-0.12	-0.75	0.5	-14	529	-0.14	-0.58	0.31
	On, LpMTG, Reading	18	539	55	541	63	N	-0.03	-0.66	0.59	-2	158			
Westwood et al. (2017), 2	On, LIFG, Related	17	669	62	694	85	N	-0.32	-0.98	0.34	-25	279	-0.33	-0.8	0.13
	On, LIFG, Mixed	17	653	66	672	88	N	-0.23	-0.88	0.42	-19	302			
Wirth et al. (2011)	On, LdlPFC, Related	20	628	67	626	72	N	0.03	-0.57	0.62	2	194	-0.04	-0.46	0.38
	On, LdlPFC, Mixed	20	584	67	589	72	N	-0.07	-0.66	0.53	-5	194			
	Off, LdlPFC, Naming	20	689	65	692	73	N	-0.04	-0.64	0.55	-3	193			

Legend: Ass = Associated; AssUn = Associated unrelated; CI =95% Confidence Intervals; CP5 = Wernicke's area; g = Hedges' g; LdlPFC = Left dorsolateral prefrontal cortex; LIFG = Left inferior frontal gyrus; LpMTG = Left posterior temporal gyrus; LSTG = Left superior temporal gyrus; M_{Sham} - M_{tDCS} = Mean difference between sham and tDCS; n = Number of participants; On = Online; Off = Offline; Sem = Semantic; SemUn = Sematic Unrelated; Sig = As reported by authors in original paper; Stro = Stroop neutral condition; V = Composite variance

Table 2 – Studies used in Primary and Moderator analyses with accuracy as the dependent variable.

								Between-participants estimate			Composite effects size estimate				
Author	**Condition**	***n***	M_{Sham}	SD_{Sham}	M_{tDCS}	SD_{Sham}	**Sig.**	***g***	CI_{Lower}	CI_{Upper}	$M_{Sham} - M_{tDCS}$	***V***	***g***	CI_{Lower}	CI_{Upper}
Fertonani et al. (2014)	Off, LdlPFC, Actions	20	6	7	5	7	N	0.1	-0.5	0.7	0.5	1	0.1	-0.3	0.5
	Off, LdlPFC, Objects	20	1	4	1	4	N	0	-0.6	0.6					
	Off, LdlPFC, Actions	20	6	7	6	7	N	0	-0.6	0.6	0	0.9	0	-0.4	0.4
	Off, LdlPFC, Objects	20	1	4	1	2	N	0	-0.6	0.6					
Henseler et al. (2014)	On, LIFG, Ass	36	2.2	0.8	2.4	3	N	-0.1	-0.5	0.4	-0.6	0.2	-0.2	-0.6	0.1
	On, LIFG, AssUn	36	2.4	0.9	3.3	3.6	N	-0.3	-0.7	0.2					
	On, LIFG, Sem	36	3.6	1.3	3.6	3.6	N	0	-0.5	0.5					
	On, LIFG, SemUn	36	2.5	0.7	2	3	N	0.2	-0.3	0.6					
	On, LMTG, Ass	36	*2.2*	*3*	2.6	2.4	N	-0.1	-0.6	0.3					
	On, LMTG, AssUn	36	*2.4*	*3*	2.2	3	N	0.1	-0.4	0.5					
	On, LMTG, Sem	36	*3.6*	*4.2*	3.7	4.2	N	0	-0.4	0.5					
	On, LMTG, SemUn	36	*2.5*	*3*	2.8	3	N	0.1	-0.4	0.6					
Jeon and Han (2012)	Off, LdlPFC, BNT	8	12	10	6.6	6.33	N	0.6	-0.4	1.5	5.24	17.1	0.6	-0.4	1.5
Ross et al. (2010)	On, LATL, Faces	15	73	19	70	28	N	0.1	-0.6	0.8	0	23.8	0	-0.5	0.5
	On, LATL, Places	15	68	22	71	22	N	-0.1	-0.8	0.6					
Turkeltaub et al. (2012)	Off, LpTC, Reading	25	97.5	9.8	100.7	9.2	Y	0.3	-0.2	0.9	1.15	1.6	0.2	-0.2	0.6
	Off, LpTC, Decoding	25	92.8	9.6	94.4	7.8	N	0.2	-0.4	0.7					
	Off, LpTC, Word ID	25	100.4	5.6	99.8	5.6	N	-0.1	-0.6	0.4					
	Off, LpTC, Attack	25	99.6	9.7	100	7.5	N	0	-0.5	0.6					
Westwood et al. (2017), 1A	On, LIFG, Naming	18	15	8.9	16	7.2	N	-0.1	-0.7	0.5	-1	3	-0.1	-0.6	0.3
Westwood et al. (2017), 1B	On, LIFG, Naming	20	13	7	12	4.4	N	0.2	-0.4	0.8	1	1.7	0.2	-0.3	0.6
Westwood et al. (2017), 1C	On, LpMTG, Naming	18	10	6	11	5	N	-0.2	-0.8	0.5	-1	1.4	-0.2	-0.6	0.3

Legend: BNT = Boston Naming Task

Table 3 – Studies used in Semantic Interference analysis using reaction times as the dependent measure.

								Between-participants estimate			Composite effects size estimate				
Author	**Condition**	***n***	M_{Sham}	SD_{Sham}	M_{tDCS}	SD_{Sham}	**Sig.**	***g***	CI_{Lower}	CI_{Upper}	$M_{Sham} - M_{tDCS}$	***V***	***g***	CI_{Lower}	CI_{Upper}
Henseler et al. (2014)	On, LIFG	36	33	36	30	42	N	0.08	-0.38	0.53	3	35	0.08	-0.24	0.4
	On, LMTG	36	33	36	37	42	N	-0.11	-0.55	0.35	-4	35	-0.11	-0.43	0.21
Meinzer et al. (2016)	On, LIFG	24	37	34	28	28	N	0.28	-0.28	0.84	9	33	0.31	-0.09	0.7
	On, STG	24	37	34	17	36	Y	0.55	-0.04	1.14	20	41	0.62	0.19	1.04
Pisoni et al. (2012), 1	Off, LSTG	12	13	14	38	14	Y	-1.66	-2.86	-0.46	-25	13	-1.85	-2.76	-0.95
Pisoni et al. (2012), 2	Off, LIFG	12	29	21	9	17	Y	0.97	0.05	1.9	20	25	1.07	0.39	1.75
Westwood et al. (2017), 1A	On, LIFG	18	72	38	95	59	N	-0.44	-1.1	0.21	-23	124	-0.46	-0.93	0
Westwood et al. (2017), 1B	On, LIFG	20	76	58	57	56	N	0.32	-0.29	0.93	19	130	0.36	-0.08	0.79
Westwood et al. (2017), 1C	On, LpMTG	18	48	80	41	70	N	0.09	-0.54	0.71	7	254	0.1	-0.34	0.54
Westwood et al. (2017), 2	On, LIFG	17	54	85	92	69	N	-0.47	-1.15	0.21	-38	291	-0.51	-1	-0.03
Wirth et al. (2011)	On, LdlPFC	20	44	11	37	12	Y	0.58	-0.06	1.23	7	5	0.65	0.19	1.12

Table 4 – Studies used in Secondary Analysis with reaction times and accuracy as dependent variables.

								Between-participants estimate			Composite effects size estimate				
Author	**Condition**	***n***	M_{Sham}	SD_{Sham}	M_{tDCS}	SD_{Sham}	**Sig.**	***g***	CI_{Lower}	CI_{Upper}	M_{Sham} - M_{tDCS}	***V***	***g***	CI_{Lower}	CI_{Upper}
Boehringer et al. (2013)	C, Off, RCereb, Reading	39	13.2	1.9	13.2	1.3	N	0	-0.4	0.44	0	0.06	0	-0.31	0.31
Fertonani et al. (2010), 1	C, Off, LdlPFC, Objects	12	739	81	761	84	N	0.25	-0.5	1.01	15.5	547	0.18	-0.35	0.72
	C, Off, LdlPFC, Actions	12	907	104	916	129	N	0.07	-0.7	0.82					
Fertonani et al. (2010), 2	C, Off, LdlPFC, Objects	12	617	51	616	51	N	-0.02	-0.8	0.73	-16.5	268	-0.28	-0.82	0.26
	C, Off, LdlPFC, Actions	12	789	100	757	60	N	-0.36	-1.1	0.41					
Jeon and Han (2012)	A, Off, RdlPFC, BNT	8	5.57	4.33	5.42	7.22	Y	-0.02	-1	0.9	-0.15	9	-0.02	-0.95	0.9
Jeon and Han (2012)	A, Off, RdlPFC, Stro	8	10	4	9	3	N	-0.27	-1.2	0.66	-1	3	-0.02	-0.95	0.9
Pope and Miall (2012)	A, Off, RCereb, Nouns	22	0.52	0.07	0.48	0.04	N	-0.69	-1.3	-0.09	-0.04	0.0003	-0.68	-1.28	-0.09
	A, Off, RCereb, Verbs	22	0.52	0.1	0.48	0.04	N	-0.52	-1.1	0.07					
Pope and Miall (2012)	C, Off, RCereb, Nouns	22	0.52	0.07	0.44	0.05	N	1.29	0.65	1.93	0.08	0.0004	1.37	0.72	2.01
	C, Off, RCereb, Verbs	22	0.52	0.1	0.45	0.05	N	0.87	0.26	1.48					
Ross et al. (2010)	A, On, RATL, Faces	15	73	19	62	14	Y	-0.62	-1.4	0.12	0	15	0	-0.48	0.48
	A, On, RATL, Places	15	68	22	79	15	N	0.55	-0.2	1.28					
Sparing et al. (2008)	C, On, CP5	15	531	62	525	62	N	-0.15	-0.8	0.53	-9	158	-0.17	-0.65	0.31
	C, Off, CP5	15	528	70	528	58	N	-0.13	-0.8	0.55					
	C, Off, CP5, 5mins	15	535	66	516	58	N	-0.37	-1.1	0.34					
	C, Off, CP5, 10mins	15	524	74	514	66	N	-0.26	-1	0.43					
Sparing et al. (2008)	C, On, CP6	15	531	367	522	54	N	-0.09	-0.8	0.59	-14	145	-0.29	-0.78	0.2
	C, Off, CP6	15	528	412	519	58	N	0	-0.7	0.68					
	C, Off, CP6, 5mins	15	535	390	513	46	N	-0.29	-1	0.4					
	C, Off, CP6, 10mins	15	524	435	507	46	N	-0.13	-0.8	0.55					
Younger et al. (2016)	A, Off, LIPL, Reading	11	91.7	8.2	92.2	8	Y	0.06	-0.7	0.82	7	21	0.6	-0.18	1.39
Younger et al. (2016)	A, Off, RIPL, Reading	11	91.7	8.2	98.8	14	N	0.61	-0.2	1.39	1	11	0.06	-0.71	0.82

Appendix 4.1 – summary of protocols used by previous studies.

Table 1: table summarising protocols used by previous studies measuring effects on verbal fluency and probe tasks including protocol used in present study at the bottom in bold.

Author	A,C	Timing	Target	Active cm2	mA	mA/cm2	Mins	Ref	Task	Sig?
Cattaneo et al. (2011), Exp 1	A	Off	LIFG	35	2	0.06	20	CS	PF,SF	Y
Cattaneo et al. (2011), Exp 2	A	Off	RIFG	35	2	0.06	20	CS	PF,SF	N
Cerruti & Schlaug (2009), Exp 1	A,C	Off	LdlPFC	16	1	0.06	20	CS	SF	N
Cerruti & Schlaug (2009), Exp 2	A	Off	R/LdlPFC	16	1	0.06	20	CS	SF	N
Ehlis et al. (2016)	A,C	Off	LIFG	35	1	0.03	20	CS	PF,SF	N
Martin et al. (2017)	A	On	M1	35	1	0.03	30	CS/RM	SF	Y
Meinzer et al. (2012)	A	On	LIFG+ATL	35	1	0.03	17	CS	SF	Y
Penolazzi et al. (2013)	A	Off(+20)	LIFG	35	2	0.06	20	CS	SF	Y
	A	Off(+20)	LIFG+ATL	35	2	0.06	20	CS	SF	N
	A	Off(+20)	LIFG+ATL	35	2	0.06	20	RH	SF	N
	A	Off(+20)	LIFG+ATL	35	2	0.06	20	CS	SF	N
Pisoni et al. (2017)	A	On	LIFG	16	0.75	0.05	20	CS	PF,SF	Y
Vannorsdall et al. (2012)	A	On	LdlPFC	25	1	0.04	30	V	PF,SF	N
Vannorsdall et al. (2016)	A	Off	LIFG	35	2	0.06	20	CS	PF,SF	Y
Boggio et al. (2009)	A	On/Off	LTC	35	2	0.06	10	RH	DRM	Y
Diez et al. (2017)	A,C	On/Off	LATL	35	2	0.06	20	RS	DRM	Y
Ferrucci et al. (2008)	A,C	On	RC	21	2	0.1	15	RD	S	Y
Ferrucci et al. (2008)	A,C	On	LdlPFC	21	2	0.1	15	RD	S	Y
Gladwin et al. (2012)	A	On/Off	LDLPFC	35	1	0.03	10	CS	MS	Y
Marshal et al. (2005)	A,C	On	L/RdlPFC	0.8	0.26	0.33	15s*	M	S	Y
Mulquiney et al. (2011)	A	Online	LdlPFC	35	1	0.03	10	CS	S	N
Pergolizzi et al. (2015)	A	Offline	LPC	35	2	0.06	10	RH	DRM	Y
Pergolizzi et al. (2015)	A	On/Off	LPC	35	2	0.06	20	RH	MS	Y
Pisoni et al. (2014)	A	On	LPPC	35	1.5	0.04	15	RH	S	Y
Pisoni et al. (2014)	A	On	LTC	35	1.5	0.04	15	RH	S	Y

Teo et al. (2011)	A	Off	LdlPFC	35	1	0.03/0.06	20	CS	S	N
Teo et al. (2011)	A	Off	LdlPFC	35	2	0.03/0.06	20	CS	S	N
Present study	A,S	On	LIFG	25	1.5	0.06	25	CS	-	-

Legend: A = anodal; ATL = anterior temporal lobe; C = cathodal; CS = contralateral supraorbital area; DRM = Deese-Roediger-McDermott; IFG = inferior frontal gyrus; L = left; M = mastoid; M1 = primary motor cortex; MS = modified Sternberg; Off = offline; Off(+20) = offline 20mins after stimulation cessation; On = online; PF = phonemic fluency; PPC = posterior parietal cortex; dlPFC = dorsolateral prefrontal cortex; T = temporal lobe; LPC = parietal cortex; R = right; RC = right cerebellum; RD= right deltoid; RH = right homologue; S = Sternberg task; SF = semantic fluency; V = vertex

Appendix 4.2 – Lexical variables (word length and frequency, CELEX Database, Baayen et al., 1995) used in all versions of the recent-probe and semantic probe task.

Table 2: lexical variables for recent-negative probe (Version A and B); Mean (Standard Deviation). Version A was used in Experiment 1.

Version A	N	R-N	N-R-N	P
List items				
Length	4 (1)	4 (1)	4 (1)	4 (1)
Probes				
Frequency	191 (335)	282 (766)	122 (217)	91 (144)
Length	4 (1)	4 (1)	4 (1)	4 (1)
Version B				
List items				
Length	4 (1)	4 (1)	4 (1)	4 (1)
Probes				
Frequency	72 (84)	80 (130)	66 (82)	50 (60)
Length	4 (1)	4 (1)	4 (1)	4 (1)

Legend: N = negative; R-N = recent-negative; N-R-N = non-recent-negative; P = positive

Table 3: lexical variables for semantic-associated probe version used in Experiment 1; Mean (Standard Deviation)

Version A	N-A	N-C	N-U	P-R	P-U
List items					
Length	5.7 (.4)	5.6 (.3)	5.5 (.3)	5.4 (.3)	5.6 (.1)
Probes					
Frequency	52 (85)	26 (23)	22 (24)	55 (62)	67 (117)
Length	5.6 (2)	5.1 (1)	5.5 (1)	5.4 (2)	5.8 (1.5)

Legend: N-A = negative-associated; N-C = negative-combined; N-U = negative-unrelated; P-R = positive-related; P-U = positive-unrelated

Table 4: lexical variables for semantic-associated probe (Version A and B); Mean (Standard Deviation)

Version A	N-A/C	N-A	N-C	N-U	P-R	P-U
List items						
Length	5.4 (.5)	5.8 (.3)	5.7 (.6)	5.5 (.2)	5.2 (.3)	5.7 (.2)
Probe						
Frequency	59 (112)	55 (112)	24 (20)	25 (26)	47 (47)	67 (124)
Length	5.7 (1.01)	5.5 (2)	5.6 (1)	5.3 (1)	5.1 (1)	5.8 (1)
Version B						
List items						
Length	5.7 (.7)	5.5 (.6)	5.5 (.1)	5.3 (.3)	5.7 (.2)	5.5 (.3)
Probe						

Frequency	55 (28)	50 (53)	29 (27)	27 (41)	51 (66)	61 (94)
Length	5.8 (.2)	5.6 (2)	4.5 (1)	6 (1)	5.8 (2)	5.6 (1)

Legend: N-A/C = negative-associated plus combined; N-A = negative-associated; N-C = negative-combined; N-U = negative-unrelated; P-R = positive-related; P-U = positive-unrelated

Appendix 4.3 – Scoring rules for clustering and switching based on Troyer et al., 1997; see also Troyer & Moscovitch, 2006 in Poreh, 2006) copied verbatim and our own criterion (underlined).

For each protocol, six scores were calculated, including the total number of correct words generated, mean cluster size, and number of switches for phonemic and semantic fluency, respectively. These scores are defined as follows:

Total number of correct words generated. This was calculated as the sum of all words produced, excluding errors and repetitions.

Mean cluster size. Cluster size was counted starting with the second word in a cluster. That is, a single word was given a cluster size of 0, two words had a cluster size of 1, three words had a cluster size of 2, and so forth. Errors and repetitions were included. The mean cluster size was computed across the three phonemic trials and across the one or two semantic trials.

Number of switches. This was calculated as the total number of transitions between clusters, in- cluding single words, for the three phonemic trials combined and for the one or two semantic trials combined. Errors and repetitions were included.

Phonemic Fluency
Clusters on phonemic fluency trials consisted of successively generated words which shared any of the following phonemic characteristics:

First letters: words beginning with same first two letters, such as ''arm'' and ''art''
Rhymes: words that rhyme, such as ''sand'' and ''stand''
First and last sounds: words differing only by a vowel sound, regardless of the actual spelling, such as ''sat,'' ''seat,'' ''soot,'' ''sight,'' and ''sought''
Homonyms: words with two or more different spellings, such as ''some'' and ''sum,'' as indicated by the participant

Semantic Fluency
Clusters on semantic fluency trials consisted of successively generated words belonging to the same subcategories, as specified below. Commonly generated examples are listed for each subcategory, although listings are not exhaustive.

Animals

Living Environment

Africa: aardvark, antelope, buffalo, camel, chameleon, cheetah, chimpanzee, cobra, eland, elephant, gazelle, giraffe, gnu, gorilla, hippopotamus, hyena, impala, jackal, lemur, leopard, lion, manatee, mongoose, monkey, ostrich, panther, rhinoceros, tiger, wildebeest, warthog, zebra, meerkat
Australian animals: emu, kangaroo, kiwi, opossum, platypus, Tasmanian devil, wallaby, wombat

Arctic/Far North animals: auk, caribou, musk ox, penguin, polar bear, reindeer, seal.
Farm animals: chicken, cow, donkey, ferret, goat, horse, mule, pig, sheep, turkey, duck, owl
North America animals: badger, bear, beaver, bobcat, caribou, chipmunk, cougar, deer, elk, fox, moose, mountain lion, puma, rabbit, raccoon, skunk, squirrel, wolf
Water animals: alligator, auk, beaver, crocodile, dolphin, fish, frog, lobster, manatee, muskrat, newt, octopus, otter, oyster, penguin, platypus, salamander, sea lion, seal, shark, toad, turtle, whale
Woodland: badger, fox, hedgehog

Human Use

Beasts of burden: camel, donkey, horse, llama, ox
Animals used for their fur: beaver, chinchilla, fox, mink, rabbit
Pets: budgie, canary, cat, dog, gerbil, golden retriever, guinea pig, hamster, parrot, rabbit

Zoological Gardens

Birds: budgie, condor, eagle, finch, kiwi, macaw, parrot, parakeet, pelican, penguin, robin, toucan, woodpecker
Bovine: bison, buffalo, cow, musk ox, yak
Canine: coyote, dog, fox, hyena, jackal, wolf
Deers: antelope, caribou, eland, elk, gazelle, gnu, impala, moose, reindeer, wildebeest
Feline: bobcat, cat, cheetah, cougar, jaguar, leopard, lion, lynx, mountain lion, ocelot, panther, puma, tiger
Fish: bass, guppy, salmon, trout *Insects*: ant, beetle, cockroach, flea, fly, praying mantis
Insectivores: aardvark, anteater, hedgehog, mole, shrew
Primates: ape, baboon, chimpanzee, gibbon, gorilla, human, lemur, marmoset, monkey, orangutan, shrew
Rabbits: coney, hare, pika, rabbit
Reptiles/Amphibians: alligator, chameleon, crocodile, frog, gecko, iguana, lizard, newt, salamander, snake, toad, tortoise, turtle
Rodents: beaver, chinchilla, chipmunk, gerbil, gopher, groundhog, guinea pig, hamster, hedgehog,
marmot, mole, mouse, muskrat, porcupine, rat, squirrel, woodchuck
Weasels: badger, ferret, marten, mink, mongoose, otter, polecat, skunk

Supermarket Items

Fruits: applesauce, bananas, cranberries, juice, mango, nectarines, peaches, raisins
Vegetables: avocado, beans, carrots, eggplant, olives, pickles, tomatoes, zucchini
Dairy case items: cheese, cream, cream cheese, eggs, milk, sour cream, yogurt
Meats: bacon, chicken, fish, hamburger, hot dogs, pork, salmon, sausage, tuna
Beverages: coffee, juice, lemonade, milk, orange juice, pop, tea, water, wine
Condiments: jelly, ketchup, marmalade, mayonnaise, pickles, relish, salad dressing
Flavourings: chives, cinnamon, parsley, pepper, sage, salt, vanilla, vinegar
Sweets and snacks: candy, cake, crackers, donuts, gum, ice cream, pie, popcorn, pudding, torte

Grain products: barley, bread, cereal, corn meal, flour, macaroni, meal, muffins, oats, rice
Baking supplies: baking powder, cornstarch, eggs, flour, salt, shortening, spices, vanilla
Specific meals/dishes: coffee, eggs, syrup, waffles; spaghetti, tomato sauce; lettuce, onions, radishes, salad dressing; pork, beans
Household goods: ammonia, bicarb, detergent, disinfectant, gift wrap, Kleenex, magazines, mop, pans, paper bags, paper towels, stamps, tin foil, toilet paper, washing powder, washing liquid, wax paper,
Personal toiletries: aspirin, comb, deodorant, medicine, mouthwash, toothpaste, vitamins
Infrastructure: aisles, basket, butcher, cash register, cashier, grocery bags, pharmacy, price tags, shelves, shopping cart, trolleys
Essentials: bread, butter, eggs, milk
Sunday roast: beef, chicken, gravy, lamb, meat, peas, pork, stock
Electronics: computers, earphones, headphones, laptops, phones, iPods, speakers
Utensils: fork, knives, peelers, rolling pin, spatula, spoon
Clothes: hats, socks
Condiments: brown sauce, ketchup, mayonnaise, Worcester sauce

Fruits

Berries & Currants: blackcurrants, blackberries, blueberries, gooseberries, strawberries, raspberries, redcurrants
Bowl fruits: apple, banana, pears, grapes, oranges, plums, peaches
Exotic fruits: kiwi, pineapple, tomato, mango, pomegranate, cherries, melon, peach, plums, nectarine, grape fruit, papaya, avocado, figs, apricots, quinoa, coconut
Citric: blood orange, grape fruit, kumquat, lemon, limes, nectarine, oranges, tangerine
Dried fruit: dates, figs, prunes, raisins

Musical Instruments

Band: guitar, drums, keyboard, bass
Brass: bugle, flute, French horn, horn, saxophone, trombone, trumpet, tuba,
Percussion: bass drum, bongo, castanets, chimes, cymbal, glockenspiel, gong, snare drum, spoons, tambourine, tamborim, triangle, drums, whistle, xylophone
Wind instruments: bassoon, clarinet, oboe, flute, recorder, saxophone
String: violin cello bass, banjo, double bass, guitar, harp, lyre, mandolin, ukulele

General Scoring Rules
In the case where two categories overlapped, with some items belonging to both categories, some items belonging exclusively to the first category, and some items belonging exclusively to the second category, the overlapping items were assigned to both categories. For example, for ''dog, cat, tiger, lion,'' the first two items were scored as pets, and the last three items were scored as feline. ''Cat'' was included in both the pet category and the feline category.

In the case where smaller clusters were embedded within larger ones, or two categories overlapped, but all items could correctly be assigned to a single category, only the larger, common category was used. For example, for ''sly, slit, slim, slam'' all begin with ''sl,'' but

an additional cluster was not scored for the last two words which differ only by a vowel sound.

Appendix 4.4 – BIS/BAS scale given to participants and scoring criteria to participant responses taken from Carver and White (1994; http://www.psy.miami.edu/faculty/ccarver/sclBISBAS.html)

BIS/BAS Scale

Each item of this questionnaire is a statement that a person may either agree with or disagree with. For each item, indicate how much you agree or disagree with what the item says. Please respond to all the items; do not leave any blank. Choose only one response to each statement. Please be as accurate and honest as you can be. Respond to each item as if it were the only item. That is, don't worry about being "consistent" in your responses. Choose from the following four response options:

1 = very true for me
2 = somewhat true for me
3 = somewhat false for me
4 = very false for me

1. A person's family is the most important thing in life.
2. Even if something bad is about to happen to me, I rarely experience fear or nervousness.
3. I go out of my way to get things I want.
4. When I'm doing well at something I love to keep at it.
5. I'm always willing to try something new if I think it will be fun.
6. How I dress is important to me.
7. When I get something I want, I feel excited and energized.
8. Criticism or scolding hurts me quite a bit.
9. When I want something I usually go all-out to get it.
10. I will often do things for no other reason than that they might be fun.
11. It's hard for me to find the time to do things such as get a haircut.
12. If I see a chance to get something I want I move on it right away.
13. I feel pretty worried or upset when I think or know somebody is angry at me.
14. When I see an opportunity for something I like I get excited right away.
15. I often act on the spur of the moment.
16. If I think something unpleasant is going to happen I usually get pretty "worked up."
17. I often wonder why people act the way they do.
18. When good things happen to me, it affects me strongly.
19. I feel worried when I think I have done poorly at something important.
20. I crave excitement and new sensations.
21. When I go after something I use a "no holds barred" approach.
22. I have very few fears compared to my friends.
23. It would excite me to win a contest.
24. I worry about making mistakes.

Scoring Criteria

Items other than 2 and 22 are reverse-scored.
BAS Drive: 3, 9, 12, 21
BAS Fun Seeking: 5, 10, 15, 20
BAS Reward Responsiveness: 4, 7, 14, 18, 23

BIS: 2, 8, 13, 16, 19, 22, 24
Items 1, 6, 11, 17, are fillers.

The fact that there are three BAS-related scales and only one BIS-related scales was not planned or theoretically motivated. The factors emerged empirically, from an item set that was intended to capture diverse manifestations of the BAS, according to various theoretical statements. It is likely that a broader sampling of items on the BIS side would also have resulted in more than one scale. I do not encourage combining the BAS scales, however, because they do turn out to focus on different aspects of incentive sensitivity. In particular, Fun Seeking is known to have elements of impulsiveness that are not contained in the other scales.

Appendix 4.5 – Supplementary Material

Here, we considered that the tDCS outcome may differ according to individual variation in the response to tDCS and to item difficulty. For individual variation, we anticipated that the response to tDCS may be positive in some participants but negative in others, so we considered the magnitude of the effects of tDCS regardless of the outcome direction. Thus, the absolute difference between testing sessions was compared between the experimental group (where sham and tDCS was applied) with the control group (where no stimulation was applied). This was done for overall performance and for our manipulation (e.g., switching, interference control). We expected that the difference would be greater in the experimental group if tDCS had an effect because the difference between sessions would be a combination of natural inter-session variation and the tDCS effect. For item difficulty, we anticipated that in the probe tasks, the effect of tDCS was more likely to be detected on more difficult items (see also Ross, McCoy, Wolk, Coslett, & Olson, 2010). We therefore ran a so-called Vincentisation analysis, in which we rank ordered reaction times for each participant across probe condition then separated reaction times into four bins ranked according to speed (e.g., very slow, slow, fast, very fast). This was done separately for each probe task.

4.5.1 Direction-neutral effects of stimulation

The absolute magnitude of the difference between testing sessions for each participant group and task are shown in Figure 4.6 (for verbal fluency data) and Figure 4.7 (for probe task data). We compared the absolute values of intersession differences between experiment and control participants (e.g., Sham minus Real, for the experimental group; Pseudo-Sham minus Pseudo-Real for the control group) using Mann-Whitney U tests (as values were non-normally distributed).

4.5.1.1 Effect on overall performance

Here, we used absolute differences between sessions in terms of average overall performance as the dependent measure. The ratio of participants who improved versus worsened did not significantly differ between experimental and control participants in fluency ($\chi^2(1) = .55$, $p = .46$), recent-probe ($_{RTs}\chi^2(1) = 1.47$, $p = .23$; $_{errors}\chi^2(1) = .05$, $p = .82$) and semantic-associated probe ($_{RTs}\chi^2(1) = 2.59$, $p = .11$; $_{errors}\chi^2(1) = .35$, $p = .55$). Mann-Whitney U tests

showed that in no experiment was there a significance effect of Condition, either in fluency, recent-probe, semantic-associated probe, see Figure 4.6 and 4.7.

4.5.1.2 Effect on switching and interference control

Here, we used the absolute differences between sessions in terms of switches and cluster sizes averaged across phonemic and semantic fluency and interference effects for probe tasks Again, we observed that the improve:worsen ratio did not significantly differ between experiment and control participants in fluency (${}_{\text{switches}}\chi^2(1) = .05$, $p = .82$; ${}_{\text{cluster size}}\chi^2(1) = .08$, $p = .78$), recent-probe (${}_{\text{RTs}}\chi^2(1) = .003$, $p = .96$; ${}_{\text{errors}}\chi^2(1) = .27$, $p = .60$) and semantic-associated probe (${}_{\text{RTs}}\chi^2(1) = .27$, $p = .61$; ${}_{\text{errors}}\chi^2(1) = 3.38$, $p = .07$). Mann-Whitney U tests showed that in no experiment was there a significance effect of Condition, either in recent-probe, semantic-associated probe, see Fig. 4.6 and 4.7.

4.5.2 Effects of stimulation by item difficulty

Results in Figure 4.8 show that stimulation had no observable effect in the experimental across speed bins. We carried out mixed factor ANOVAs separately for each experiment, with *Speed Bin* (Very Fast, Fast, Slow, Very Slow) and *Condition* (Sham vs Real for the experimental group; Pseudo-Sham vs Pseudo-Real for control group) as within-participants factors, and *Group* (Experiment vs Control) as a between-participants factor. We do not report effects of speed bins as these were expected and not of interest here. Crucially, there was no *Speed Bin* x *Group* x *Condition* interaction for recent-probe ($F(3,126) = .76$, $p = .52$, $\eta_p^2 = .02$) and semantic-associated probe ($F(3,125) = .64$, $p = .12$, $\eta_p^2 = .05$).

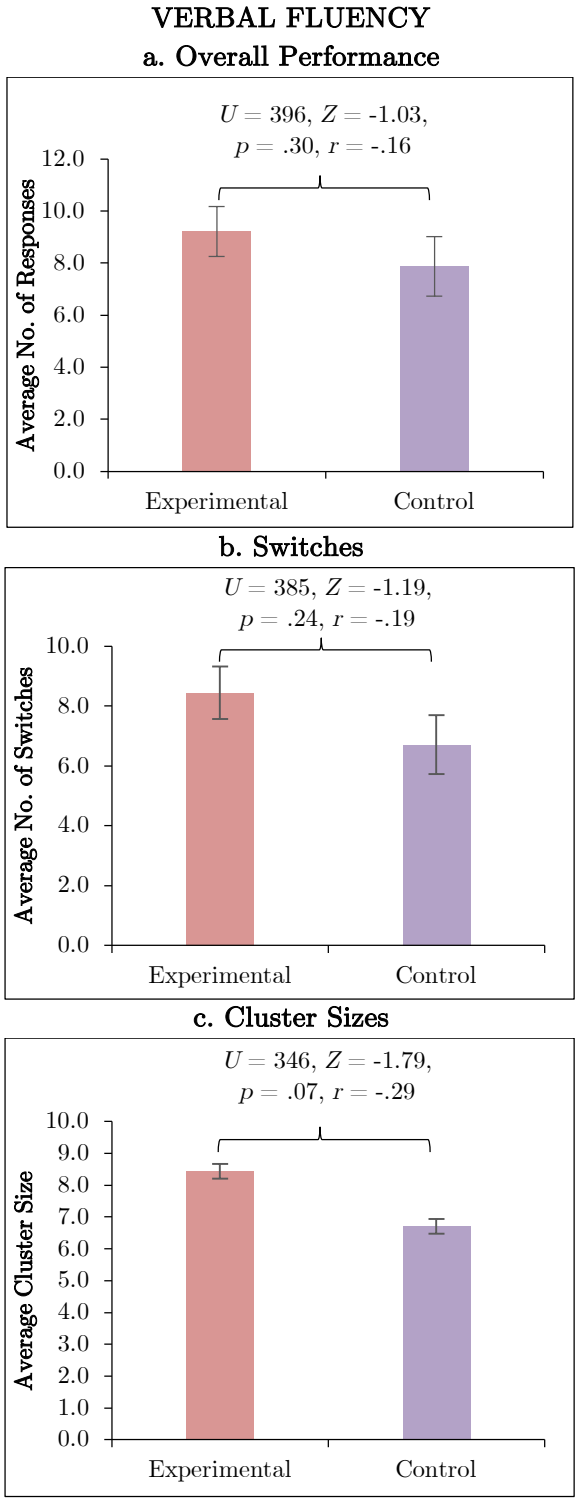
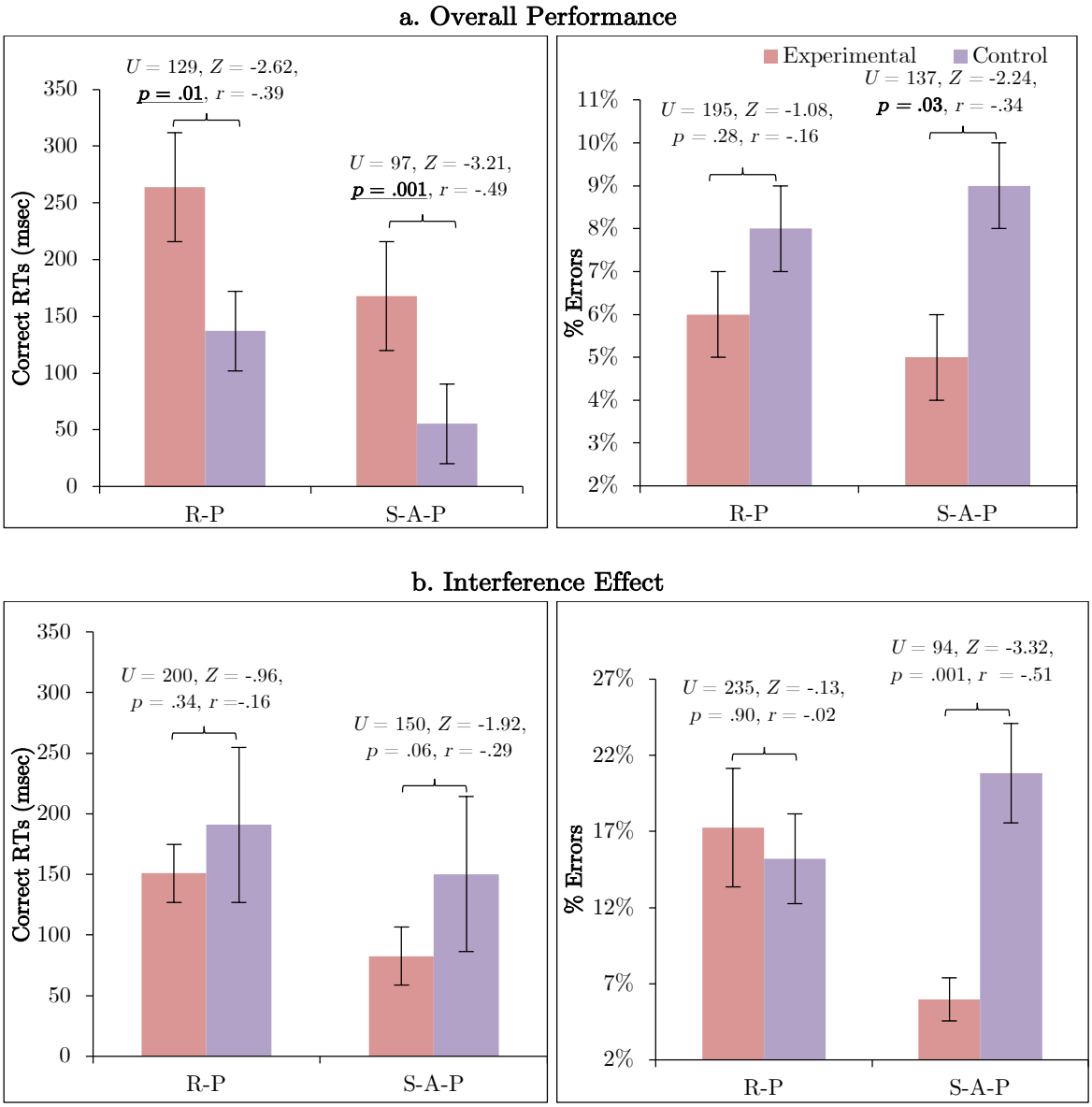


Fig. 4.6 - Absolute intersession differences across participant groups in terms of overall performance (panel a), switches (panel b), and cluster sizes (panel c). Error Bars indicate Standard Error.



Legend: R-P = Recent-Probe; S-A-P = Semantic-Associated Probe

Fig. 4.7 - Absolute intersession differences across participant groups and probe tasks, in terms of differences in reaction times and percentage errors for overall performance (panel and aggregated interference (panel b). Error Bars indicate Standard Error.

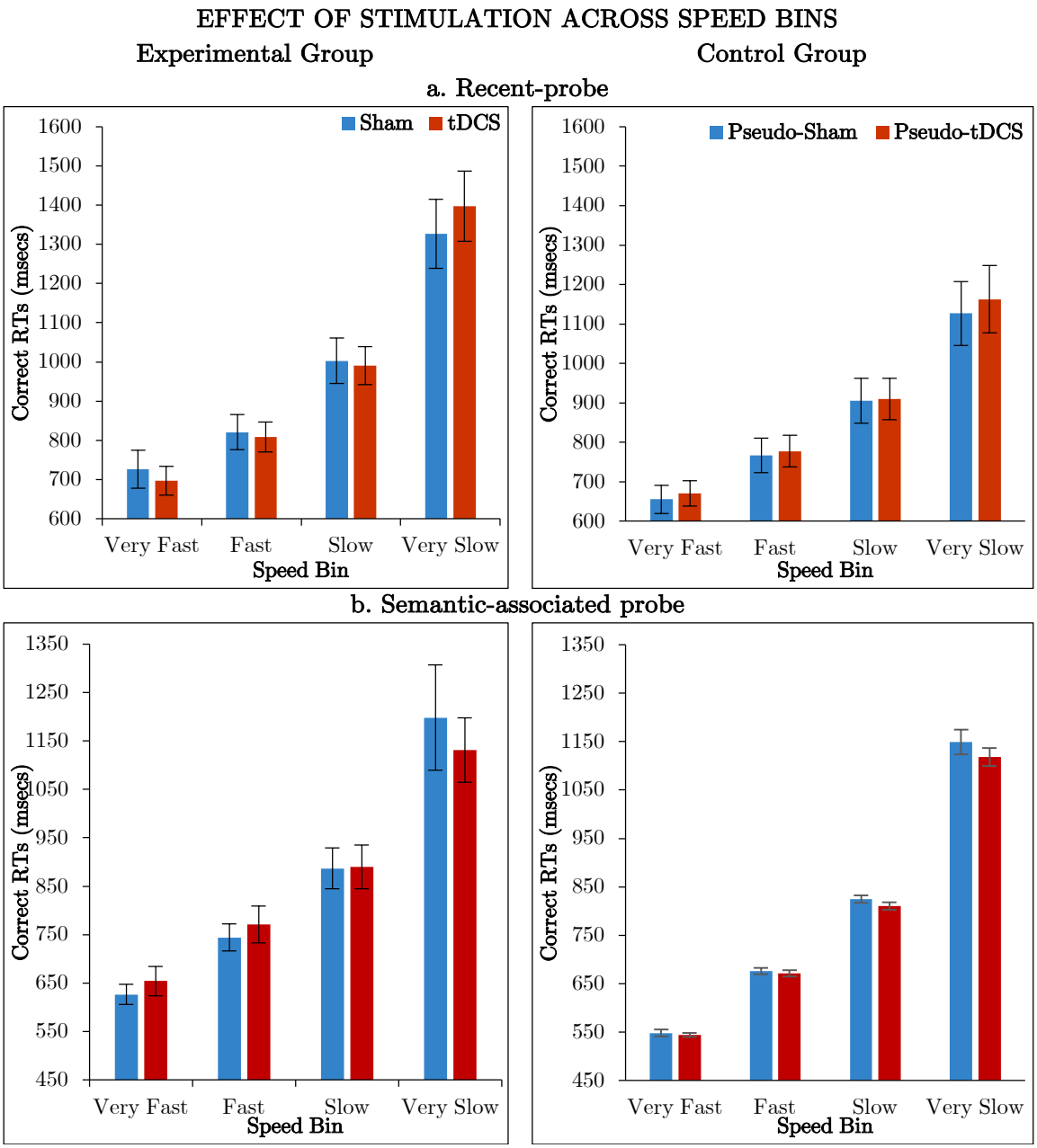


Fig. 4.8 – Average reaction times following Vincentisation across speed bins, participant groups and probe tasks. Error Bars indicate Standard Error

4.5.3 Summary

In summary, we found no effect of stimulation, either when discounting antagonistic responses to tDCS or when looking at item difficulty. Specifically, we found no evidence that absolute differences between testing sessions in the experimental (when sham and tDCS was applied) versus the control group (when stimulation was not applied). In verbal fluency, we see systematically greater variation in the experimental group compared to control, but this is only minor and in any case not statistically significant. In probe tasks, we see no systematically greater variation in the experimental group relative to the control. Variation

in the experimental group was significantly greater relative to control for overall performance in terms of reaction times, which is in line with our prediction. However, we are cautious when drawing an inference from this finding. Reponses were slower and more variable in the experimental group compared to controls (see Figure 4.8), so greater variation between sessions was likely due to natural group differences masquerading as an effect of tDCS. Moreover, when comparing overall performance with interference control, or reaction times with percentage errors, we see patterns of data that are the mirror images of each other. Finally, in terms of item difficulty, tDCS did not differentially impact on performance with respect to the difficulty of probe items. We see that tDCS reduced reaction times for all except very slow reaction times in recent-probe, whilst the opposite is true for semantic-associated probe.

Appendix 5.1 -- Sample of studies included in the review, with details on stimulation parameters and a summary of findings

Author	N	Design	A,C	Timing	mA	mA/cm^2	Anode (cm^2)	Mins	Target	Ref	Task
Cattaneo et al (2011), Exp 1	10	W	A	Off	2	0.06	35	20	LIFG	CS	P,S
Cattaneo et al (2011), Exp 2	8	W	A	Off	2	0.06	35	20	RIFG	CS	P,S
Summary: In two experiments, participants performed phonemic and semantic fluency immediately after anodal tDCS applied to the left (Experiment 1) or right (Experiment 2) left inferior frontal gyrus (LIFG). Only LIFG stimulation resulted in a significant improvement compared to sham, with this improvement seen across both fluency tasks.											
Cerruti & Schlaug (2009), Exp1	18	W	A,C	On	1	0.06	16	20	LdlPFC	CS	P
Cerruti & Schlaug (2009), Exp2	12	W	A	Off	1	0.06	16	20	R/LdlPFC	CS	P
Summary: In two experiments, participants performed a phonemic fluency task either during (Experiment 1) or after (Experiment 2) stimulation. For Experiment 1, fluency was performed before and during the last 4mins of anodal or cathodal tDCS of the left dorsolateral prefrontal cortex (dlPFC) in separate sessions spaced three hours apart. For Experiment 2, fluency was performed before and after anodal tDCS, with the left or right dlPFC being targeted in separate sessions performed on separate days. Participants performed a variant of phonemic fluency, where one had ninety seconds to generate responses. The results showed a null effect of tDCS in both experiments.											
Ehlis et al (2016)	23	W	A,C	Off	1	0.03	35	20	LIFG	CS	P,S
Summary: Participants performed phonemic and semantic fluency tasks after anodal tDCS applied to the LIFG. Participants performed a variant of a verbal fluency task in which they were given only thirty seconds for each letter and category cue. The results showed that tDCS had no effect on performance.											
Martin et al. (2017)	24	W	A	On	1	0.03	35	30	M1	CS/RM	S
Summary: Participants produced significantly fewer errors during anodal stimulation of the left primary motor cortex (of M1). Participants performed a variant of the typical semantic fluency task. This task was a paced, and participants were asked to give ten examples for six different semantic categories.											
Meinzer et al (2012)	20	W	A	On	1	0.03	35	17	LIFG+T	CS	S
Summary: Participants produced significantly more words during anodal stimulation of the left IFG (and partial anterior temporal) region. Participants performed the same self-paced semantic fluency task as Martin et al. (2017)											
Penolazzi et al, 2013	19	W	A	Off(+20)	2	0.06	35	20	LIFG	CS	S
Penolazzi et al, 2013	19	W	A	Off(+20)	2	0.06	35	20	LIFG+T	CS	S
Penolazzi et al, 2013	19	W	A	Off(+20)	2	0.06	35	20	LIFG	RH	S
Penolazzi et al, 2013	19	W	A	Off(+20)	2	0.06	35	20	LIFG	CS	S
Summary: The authors investigated the efficacy of four different montages applying anodal tDCS to the left IFG and local areas. The montages were based on those used in previous studies, and included left inferior frontal gyrus (frontal stimulation), left IFG plus the anterior temporal lobes (frontal-temporal stimulation), the left IFG plus anterior temporal lobes with cathode over the right homologue area (bilateral stimulation), or the left IFG plus the anterior temporal lobe with a cathode of bigger size (100cm2) over the right supra-orbital area (unilateral stimulation). Performance on semantic fluency was measured immediately before stimulation, and again immediately and 20mins after stimulation cessation. Analysis of correct responses based on the highly stringent scoring criteria (i.e., on the items unambiguously rated as members of a given category) showed that frontal stimulation improved fluency compared to sham, but only in 18mins after											

stimulation. This effect, however, was not significant with less stringent scoring criteria (i.e., on the items ambiguously rated as members of a given category, plus items rated as more peripheral members).

Pisoni et al. (2017)	18	W	A	On	0.75	0.05	16	20	LIFG	CS	P,S

Summary: The authors used EEG recordings to measure the response to TMS over the LIFG before and after anodal tDCS applied to this brain region. During stimulation participants performed a phonemic and semantic fluency tasks, with tDCS improving overall performance. Data was not reported separately for each fluency task.

Vannorsdall et al (2012)	12	W	A	On	1	0.04	25	30	LdlPFC	V	P,S

Summary: Participants were given anodal or cathodal tDCS applied to the LdlPFC during which they performed an object naming, oral reading task and in the last the last six minutes they performed a phonemic and semantic fluency task. The authors found no effect of stimulation in terms overall performance and switching, but they did find a small but significant effect on clustering ($p = .055$).

Vannorsdall et al (2016)	14	W	A	Off	2	0.06	35	20	LIFG	CS	P,S

Summary: The authors reported are failed attempted replication of Catteneo et al. (2011). However, Cattaneo et al. (2016) criticised this study for its notable differences from their original. For example, participants were given the same stimuli in both sessions, and sessions were spaced at least twenty-four hours apart from each other, which increase the potential for practice effects that could reduce tDCS related improvements on task performance.

Westwood et al (2017)	49	W	A	On	1.5	0.06	25	25	LIFG	CS	P,S

Summary: Participants were given anodal tDCS to the LIFG during a phonemic and semantic fluency task. The results showed no significant effect of stimulation.

Legend: A = anodal; C = cathodal; Off = offline; On = online;

Appendix 6.1 – list of stimuli including lexical variables

Set	Italian (English)	CVC	No. of Syllables	Length	Frequency	AoA
Set A	Calza (Sock)	CVCCV	2	4	3.1	2.9
	Capra (Goat)	CVCCV	2	4	12.4	5.2
	Cornice (Frame)	CVCCVCV	3	5	26.6	7.7
	Formica (Ant)	CVCCVCV	3	3	4.0	4.3
	Gufo (Owl)	CVCV	2	3	3.3	6.2
	Mare (Sea)	CVCV	2	3	168.1	4.7
	Pala (Spade)	CVCV	2	5	3.1	8.1
	Palco (Stage)	CVCVCCCV	2	5	138.8	6.3
	Palestra (Gym)	CVCVCV	3	3	4.2	6.0
	Parete (Wall)	CVCVCV	3	4	144.7	3.8
	Pavone (Peacock)	CVCCV	3	7	3.0	5.8
	Pecora (Sheep)	CVCVCV	3	5	42.2	4.3
	Pietra (Stone)	CVVCCV	3	5	95.7	4.4
	Rana (Frog)	CVCV	2	4	3.8	4.3
	Sapone (Soap)	CVCVCV	3	4	21.9	3.2
	Sentiero (Path)	CVCCVVCV	4	4	53.7	6.1
	Tende (Curtains)	CVCCV	2	8	25.9	5.0
	Vagone (Carriage)	CVCVCV	3	8	13.5	5.8
	Valigia (Suitcase)	CVCVCVV	3	8	13.1	8.2
	Volpe (Fox)	CVCCV	2	3	14.3	5.0
		Mean	2.6	4.8	39.8	5.4
		SD	0.6	1.7	53.0	1.5
Set B	Camino (Fireplace)	CVCVCV	3	9	8.8	7.4
	Corona (Crown)	CVCVCV	3	5	25.0	7.8
	Fantasma (Ghost)	CVCCVCCV	3	5	21.0	5.1
	Gamba (Leg)	CVCCV	2	3	67.5	3.0
	Ladro (Thief)	CVCCV	2	5	6.5	7.2
	Lapide (Headstone)	CVCVCV	3	9	0.7	9.4
	Libro (Book)	CVCCV	2	4	255.5	3.7
	Lumaca (Slug)	CVCVCV	3	4	2.4	6.0
	Lupo (Wolf)	CVCVCV	2	4	6.9	4.5
	Matita (Pencil)	CVCVCV	3	6	16.8	4.1
	Mela (Apple)	CVCV	2	5	18.7	4.2
	Neve (Snow)	CVCVCV	2	4	62.1	4.1
	Nido (Nest)	CVCV	2	4	14.1	5.1
	Nuvola (Cloud)	CVCVCV	3	5	32.8	3.6
	Pentola (Pot)	CVCCVCV	3	3	25.9	6.0
	Ponte (Bridge)	CVCCV	2	6	59.3	5.6
	Quaderno (Notebook)	CVVCVCCV	3	8	7.8	6.4
	Rospo (Toad)	CVCCV	2	4	3.4	6.1
	Vento (Wind)	CVCCV	2	4	117.2	3.9
	Vernice (Paint)	CVCCVCV	3	5	41.9	4.5
		Mean	2.5	5.1	39.7	5.4
		SD	0.5	1.7	58.5	1.7
Set C	Barca (Ship)	CVCCV	2	4	46.9	5.3
	Bastone (Cane)	CVCCVCV	3	4	9.6	5.7
	Bocca (Mouth)	CVCCV	2	5	142.2	3.6
	Bosco (Forest)	CVCCV	2	6	71.6	6.3
	Cartina (Map)	CVCCVCV	3	3	32.2	5.6
	Cervo (Deer)	CVCCV	2	4	12.5	5.2
	Divano (Sofa)	CVCVCV	3	4	21.6	4.5
	Finestra (Window)	CVCVCCCV	3	6	139.7	4.7
	Forziere (Chest)	CVCCVVCV	4	5	46.1	5.1
	Fune (Rope)	CVCV	2	4	33.9	5.4

Ginestra (Broom)	CVCVCCCV	3	5	7.0	5.5
Gomito (Elbow)	CVCVCV	3	5	16.6	4.8
Manzo (Beef)	CVCCV	2	4	17.5	6.6
Pane (Bread)	CVCV	2	5	78.0	3.6
Piscine (Swimming Pool)	CVCCVCV	3	13	32.2	4.2
Pomata (Cream)	CVCVCV	3	5	34.5	6.0
Scudo (Shield)	CCVCV	2	6	9.0	6.5
Tavolo (Table)	CVCVCV	3	5	214.6	4.4
Vela (Sail)	CVCV	2	4	8.7	6.5
Vetro (Glass)	CVCCV	2	5	132.1	4.5
	Mean	2.6	5.1	55.3	5.2
	SD	0.6	2.0	57.8	0.9