THE UNIVERSITY OF CHICAGO

TRANSCRANIAL DIRECT CURRENT STIMULATION AND EPISODIC MEMORY RETRIEVAL

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

STEPHEN JOSEPH GRAY

CHICAGO, ILLINOIS

JUNE 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I want to sincerely thank my advisor, Dr. David Gallo, for all of his kindness and patience with me throughout the years, his prompt and insightful feedback to drafts of this dissertation and other papers I have written, and the amount of time he spent training me to become the best scientist I can be. He has truly been the best advisor for which anybody could ask. I would also like to thank my other committee members: Dr. Daniel Casasanto, Dr. Marc Berman, and Dr. Ed Awh, for taking the time to work with me and provide me with helpful ideas that directly shaped the trajectory of this project.

I would like to also thank my wife, Jacqueline, for putting up with me throughout graduate school and following me to Chicago at the potential cost of her own career development. I want to thank my mother and father, Cheryl and Kevin, and my brother, Michael, for all of their continuous support in everything that I do. I would like to also acknowledge my cats, Misty and Pixel, who continuously provided me with comfort in stressful times and forced me to work by sitting on my lap constantly throughout the writing process.

Additionally, thank you to Geoff Brookshire for all of his help in teaching me how to administer tDCS and to Daniel Casasanto for allowing me to use his lab space for the entirety of my first tDCS experiment. Also, thanks to Heather Mangelsdorf for helping proofread the dissertation and providing useful feedback.

Finally, I would like to thank all of my research assistants for all of their hard work throughout these projects. Thank you to Parina Lalchandani, George Abitante, Vannia Puig, Ran Yan, and Morgan Dufaut for all of their hours running participants and coding data – I could not have completed this research without you.

ABSTRACT

Neuroimaging and brain damage studies suggest that dorsolateral prefrontal cortex (dlPFC) is associated with cognitive control of episodic recollection. Brain stimulation studies can provide more direct evidence. If dlPFC is causally involved in retrieval, then transcranial direct current stimulation (tDCS) of this brain region should increase recollection accuracy, especially when recollection is difficult and requires high levels of cognitive control. Here, we report the first series of brain stimulation experiments to directly test this hypothesis. In four experiments, we administered tDCS to dlPFC immediately after studying to-be-learned material but just prior to recollection testing, thereby targeting retrieval processes. In Experiment 1, we found that stimulation of dlPFC significantly increased recollection accuracy, relative to a no-stimulation sham condition and to a control, parietal stimulation condition. These dlPFC stimulation effects were behaviorally selective, increasing accuracy only when participants needed to recollect difficult information. In Experiments 2 and 3, we manipulated the difficulty of the stimuli to be remembered, but failed to replicate the effects of stimulation in our whole sample analyses. However, we did find post hoc evidence of tDCS effects in morning participants. In Experiment 4 we attempted to replicate this morning effect, but we were again unable to find effects of dlPFC stimulation on recollection accuracy, although post hoc analyses revealed that early morning participants were in the predicted direction. We also failed to find a stimulation effect in an eyewitness identification paradigm. Taken together, these results indicate that the effect of tDCS over dlPFC on recollection accuracy is not very robust. These results argue against the dominant theory that dlPFC is causally involved in retrieval, or they at least indicate that tDCS is ineffective at increasing dlPFC's contribution to the retrieval process.

INTRODUCTION

One evening in 1975, Dr. Donald Thomson was discussing the psychology of eyewitness testimony on a television program (Thomson, 1988). The next day, he was brought into a local police station and accused of raping a woman and leaving her unconscious in her apartment. Fortunately for Dr. Thomson, he had the perfect alibi – he was still on television giving his talk during the exact time of the attack. The victim had been watching Dr. Thomson's talk about eyewitness testimony on television just before the attack, and had ironically confused his face with that of her attacker. This example may seem farfetched, but there are many cases in which the falsely accused are not so lucky and end up behind bars. Ronald Cotton was charged for rape in a similar circumstance and spent 11 years in prison before DNA testing revealed he was not the attacker (Thompson-Cannino, Cotton, & Torneo, 2009). These errors in eyewitness identification are considered false memories, or distortions in the details of our memories about things that never happened or happened in a different context than we remember. While some of these distortions may have drastic consequences like in the case of Ronald Cotton, even small misgivings about memory details can influence our perception of the past.

By using the tools of modern neuroscience, we can better understand and potentially enhance the brain's ability to accurately remember the past, thereby minimizing problems such as those described above. More specifically, we can explore how the brain implements retrieval monitoring: the strategies used to prevent the formation of false memories during retrieval (Mitchell & Johnson, 2009). Retrieval monitoring can involve both memory search and evaluation (i.e. postretrieval monitoring) processes. The experiments in this dissertation used transcranial direct current stimulation (tDCS), a technology which can selectively increase brain function in a specific region, to modulate the function of dorsolateral prefrontal cortex (dlPFC), a

region particularly key to retrieval monitoring, just before retrieval in several memory tests. In doing so, we sought to provide casual evidence of this region's importance for retrieval monitoring. In Experiment 1, we applied anodal tDCS to left and right dlPFC just before retrieval on a memory test to see if this would increase task performance. In Experiments 2 and 3, we manipulated the difficulty of the encoding phase of the memory test to see if tDCS over dlPFC before retrieval would have selective effects based on monitoring demands. In Experiment 4, we attempted to replicate the results of Experiment 1 and to extend them to an eyewitness identification paradigm. The research presented in this dissertation represents the most comprehensive test to date of dlPFC function during retrieval through the use of modern brain stimulation technology.

False Memories and Retrieval Monitoring

False memory research has been a popular topic among psychologists for the past few decades, and multiple mechanisms explaining false memory have been proposed (Reyna & Lloyd, 1997). For example, one explanation of false memory is the activation-monitoring framework (Roediger & McDermott, 2000; Roediger, Watson, McDermott, & Gallo, 2001), which was originally proposed to explain the formation of false memories in the Deese-Roediger McDermott memory paradigm (DRM; Roediger & McDermott, 1995), and was more recently expanded to explain all false memories (Gallo, 2010). According to this expanded model, activation is conceptualized as the mental generation of misleading information. It can involve both top-down processes, such as associations, gist, inferences, categories, or schemas, and bottom-up processes, such as feature overlap, feelings of familiarity, and partial recollections. In other words, activation is generating "the false memory signal." In contrast to activation, monitoring describes any memory editing or decision process that helps clarify the source of

mentally activated information. This includes criteria-based strategies based on comparing memorial evidence to some kind of retrieval expectations, as well as corroboration-based strategies that use knowledge of collateral or co-occurring events. Monitoring strategies can be used to clarify the origin of the memory signal generated by activation and to prevent a person from accepting a false memory as true.

The idea that failures in monitoring can lead to false memories in everyday life is central to the claims of the source monitoring paradigm (see Johnson, 2006 for a review), which argues false memories are a result of a failure to distinguish between real events and imagined ones. The idea of source monitoring has been heavily influenced by dual-process theories, which differentiate between feelings of familiarity or "oldness" elicited by a retrieval cue and the specific "re-experiencing" of specific contextual details from the past (for a review of dual process theories, see Yonelinas, 2002). When a memory is recalled, it does not necessarily have a label indicating the source from whence it came. People must make decisions about whether an item in memory occurred in reality based on the "perceptual, contextual, semantic, and emotional" details of the memory itself. If a particular memory trace has a lot of specific perceptual details associated with it, then one should be more likely to conclude that the event really happened than an event with less of these details. For example, if you are trying to determine whether you really met New England Patriots quarterback Tom Brady, then remembering specific details, such as the color of the shoes he was wearing, the specific location and time of day, and the smell of his cologne (all things that you are unlikely to imagine during a daydream) may provide compelling memorial evidence that the event really happened.

Retrieval monitoring can be further broken down into two distinct types of decision processes: diagnostic monitoring and disqualifying monitoring (Gallo, 2004). Diagnostic

monitoring describes the use of retrieval expectations to "diagnose" whether one has enough evidence to determine whether an item in memory actually occurred. For an example of a diagnostic monitoring decision, consider that you are at the grocery store and trying to remember what you need to buy. Passing by birthday candles, you think to yourself, "Well, if I needed to buy birthday candles, I would definitely remember that! Since I do not, I do not need to buy them." Disqualifying monitoring describes the use of some kind of external knowledge to logically rule out the possibility of an event occurring. When you pass by milk in the grocery store, an example of a disqualifying monitoring decision might be, "Well, I had cereal this morning and did not run out of milk, so therefore I must have milk at home and do not need to buy it."

Many researchers have studied retrieval monitoring by using basic source memory tasks in which participants study items in two different sources and are asked later to identify in which of the sources a given item was originally studied. These tasks allow participants to adopt strategies in which they can reject seeing an item in a given source if they happen to remember seeing the item in the other source, also known as recall-to-reject strategies (Gallo, 2004). Thus, the extent of the claims these studies can make can only be about disqualifying monitoring decisions, which allow for the use of such logic. To study diagnostic monitoring processes, which are more akin to the types of decision processes proposed in the original source monitoring paradigm (Johnson, Hashtroudi, & Lindsay, 1993), researchers must use a task that does not allow participants to use recall-to-reject strategies, or at least controls for them.

The criterial recollection task (Gallo, Weiss, & Schacter, 2004) is a variation of traditional source memory tasks that allows researchers to measure recollection accuracy for specific kinds of information. In addition to controlling different retrieval strategies, the criterial

recollection task also has the advantage of (1) controlling for familiarity-based responding, isolating recollection effects and (2) separately assessing recollection accuracy for different kinds of materials across test blocks and within-subjects. These advantages allow for more powerful designs that are well suited for neuroimaging and brain stimulation studies. Over the past 10 years, a growing number of studies have used the criterial recollection task to study how retrieval expectations can influence memory, the effects of aging on retrieval expectations and metamemory, and the brain activity associated with retrieval monitoring (see Gallo, 2013 for a review). The studies in this dissertation use the criterial recollection task as a testbed for studying effects of brain stimulation on the recollection of different kinds of materials.

Biological Correlates of Retrieval

The neural network associated with episodic memory tasks has been studied extensively with high resolution neuroimaging techniques (e.g., Gabrieli, 1998; Shallice et al., 1994; Ranganath & Ritchey, 2012). During encoding, left prefrontal cortex (PFC) and medial temporal regions, including the hippocampus, parahippocampal cortex, and perirhinal cortex, are thought to be involved in the binding of source information to be later recollected (Mitchell & Johnson, 2009; Wagner et al., 1998). During episodic retrieval, a recollection network containing regions such as the medial temporal lobe, lateral parietal cortices, and medial PFC, is activated (for review, see Rugg & Vilberg, 2013). Because the retrieval cue rarely is a perfect match with the memory representation, the hippocampus (specifically, the CA1 subfield) is thought to partake in a pattern completion process at retrieval that compares the representation presented by the cue to the representation of the trace (Bakker, Kirwan, Miller, & Stark, 2008; Lacy et al., 2011).

Executive control of recollection is thought to rely on activity in the PFC, which effectively "directs" the reactivation of hippocampal or neocortical stored representations.

Several functional magnetic resonance imaging (fMRI) studies suggest executive control of retrieval is dependent on left anterior ventrolateral PFC (Ranganath, Johnson, & D'Esposito, 2000; Dobbins, Foley, Schacter, & Wagner, 2002), a region also typically reported to be critical for episodic memory encoding (Buckner, Wheeler, & Sheridan, 2001; Clark & Wagner, 2003). This research is highly consistent with the notion that there is significant overlap between encoding and retrieval processes (Rugg, Johnson, Park, & Uncaper, 2008; Cansino, Maquet, Dolan, & Rugg, 2002; Buckner & Koutstaal, 1998), and that retrieval is initiated by activity in left inferior, anterior regions of PFC.

The laterality of the prefrontal regions in retrieval monitoring is unclear. Several theories suggest the different hemispheres subserve different aspects of retrieval monitoring. One hypothesis is that activation of left dlPFC occurs when high levels of control or systematic source judgments are needed (Nolde, Johnson, & Raye, 1998; Velanova et al., 2003; Wheeler & Buchner, 2003; Mitchell et al., 2008), whereas right dlPFC regions subserve heuristic source monitoring processes dependent on feelings of familiarity (Cruse & Wilding, 2009; Hayama & Rugg, 2009; Rugg, Henson, & Robb, 2003). This account would predict that disqualifying monitoring processes, which involve the deliberate use of logic to inform memory decisions, would be more likely to rely on left dlPFC, whereas diagnostic monitoring, which involves the use of heuristic judgments based on some sort of retrieval expectations, would rely on right dlPFC. The *production-monitoring* hypothesis proposes that left PFC is involved in the "search" process, whereas right PFC is involved in the monitoring of retrieved, less-differentiated information (Cabeza, Locantore, & Anderson, 2003). This hypothesis does not directly conflict with the systematic-heuristic distinction so much as it simply restricts the role of left PFC to *only* the memory search process, and also emphasizes that right PFC may become involved after an

initial search attempt fails (Lepage, 2004; Henson et al., 1999). For disqualifying and diagnostic monitoring, this point of view would predict that right dlPFC is primarily responsible for both processes.

The left systematic/right heuristic account has been partially supported by a few fMRI studies demonstrating the importance of right dlPFC during the criterial recollection test for non-distinctive details about stimuli like font color (Gallo, Kensinger, & Schacter, 2006; McDonough, Wong, & Gallo, 2012). Gallo, McDonough, and Scimeca (2010) found an increase in right dlPFC activation for monitoring word stimuli in comparison to pictures, but additionally, found increased activation in left dlPFC when looking at activity from a recall-to-reject disqualifying monitoring memory task. This neural finding suggests participants were using diagnostic monitoring to some extent during the task in which recall-to-reject logic was allowed (which is why there were no differences in right dlPFC activation in the contrast between the two), and that the additional activation of left dlPFC corresponds specifically to the use of the more logical, recall-to-reject processes. Given the difficulty in distinguishing between search and decision processes at retrieval in fMRI, however, the *production-monitoring* hypothesis cannot be ruled out in this scenario. It may be the case, for example, that the exclusion process is using production to generate the criteria on which the recall-to-reject decision is being made (leading to increased left dlPFC activity), and the word test, which presents less differentiated information, requires more monitoring (right dlPFC activity).

It is difficult to distinguish between the two laterality hypotheses using fMRI and other imaging techniques for a few reasons. Given the limited temporal resolution of fMRI, it is difficult to differentiate search and decision processes at retrieval. Event related potentials (ERPs) are more suited to distinguishing between distinct search and monitoring waveforms

(e.g., Herron & Rugg, 2003), but lack the spatial resolution of fMRI to distinguish between specific PFC regions. Furthermore, both methods produce correlational data that is subject to the limitations of correlational designs, such as not being able to establish causality.

Lesion studies allow causal inferences to be made, but they are impractical because patients with prefrontal damage have a large number of cognitive issues that prevent them from understanding (let alone completing) the task demands of source monitoring paradigms (Verfaellie, Rapcsak, Keane, & Alexander, 2004). Even if patients with prefrontal lesions could understand task instructions, it would be impossible to determine whether errors were due to encoding or retrieval failures on a memory task. Because the brain damage occurs before the experiment, it would affect both encoding and retrieval processes that take place throughout the task.

No studies to the author's knowledge applied intracranial stimulation to PFC before a memory task, as this region is not a typical location to implant an electrode in the treatment of neurological disorders. Several studies, however, have found evidence that intracranial brain stimulation of medial temporal regions can lead to improvements in declarative memory (Fell et al., 2013; Susthana et al., 2012), whereas stimulation of midbrain regions seems to lead to decreases in declarative memory (Halbig et al., 2004; Troster at al., 1998). As is the caveat with interpreting most intracranial stimulation studies, these experiments were done on small populations or single participants, and these populations typically had some sort of neurological disorder (epilepsy or Parkinson's disease) that may have interacted with these results. To provide causal evidence of PFC's role in retrieval, experimental manipulation of brain function after encoding but before retrieval is required, which can only be achieved through the use of noninvasive brain stimulation methods.

In sum, although the dlPFC is clearly associated with memory retrieval, the specific functions of dlPFC during retrieval are unclear. Moreover, the potential lateralization of these functions is unclear. Neuroimaging techniques can only provide correlational evidence about these questions. Lesion studies are impractical because of the difficulty of distinguishing whether errors are due to deficits in encoding and retrieval. The best methods to approach these questions are brain stimulation methods, such as transcranial direct current stimulation. By using converging evidence from brain stimulation and neuroimaging techniques, a clear picture of the causal relationship between a brain area and a cognitive process can emerge.

Mechanisms of Brain Stimulation

Using brain stimulation techniques to investigate the relationship between brain and behavior began 35 years ago with the introduction of Transcranial Electric Stimulation (Merton & Morton, 1980), a historical precursor of modern stimulation techniques like transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS). Merton and Morton (1980) demonstrated that a brief, high-intensity electric shock to the scalp over motor cortex could produce a corresponding muscular response. A few years later, Barker, Jalinous, and Freeston (1985) conceptually reproduced these findings by applying magnetic stimulation to the scalp over motor cortex and evoking movement on the contralateral side of the body. This was the first of many studies that would use magnetic stimulation to affect neural activity (see Hallett, 2000 for review).

Today, TMS is the most commonly used technique in noninvasive brain stimulation research and also has clinical use in measuring motor-evoked potentials to evaluate the effects of traumatic brain injuries or stroke (Groppa et al., 2012). TMS takes advantage of Faraday's Law of electromagnetism, which states that a changing or pulsing magnetic field will induce an

electric current perpendicular to its axis. By applying a controlled magnetic pulse (or repetitive pulses), TMS effectively induces currents within a large group of neurons. Depending on the parameters of the pulse (such as frequency) and the targeted area of cortex, TMS can either induce or inhibit activity in a given brain region for a short period of time.

While its ability to temporarily and noninvasively influence brain function in humans with good spatial resolution is unparalleled, TMS has several significant drawbacks. It is difficult to have a convincing control condition in TMS studies (Duecker & Sack, 2015). There are two general types of control conditions in TMS studies: sham control stimulation, in which the participant receives minimal or no stimulation as a result of either using a fake coil or turning the stimulation device to an angle such that stimulation is not intense, and off-target active stimulation, in which an area of the brain that is not relevant for the particular experiment is stimulated (Davis, Gold, Pascual-Leone, & Bracewell, 2013). Not stimulating control participants may lead to differences in task performance between groups as a result of unanticipated external factors, such as the sound of a TMS machine for one group but not the other (Duecker & Sack, 2013), or produce differences in sensations that participants may detect. Stimulating a non-relevant area of the brain may confound one's results, given the brain's high interconnectivity in general. Furthermore, although there have been significant advancements in safety protocols during the administration of TMS (Rossi et al., 2009), there are a few undesirable side effects associated with use of the technology, such as fatigue, headache, and nausea (Poreisz, Boros, Antal, & Paulus, 2007).

tDCS gives up some of the spatial resolution provided by TMS in exchange for more powerful control conditions, less pervasive side effects, and a significantly lower cost to administer. Stimulating with anodal tDCS over a region of the brain is thought to increase the

cortical excitability of that region, whereas placing a cathodal electrode over a region of the brain is thought to decrease cortical excitability of that region (Nitsche et al., 2008). The mechanisms of how tDCS influences brain activity are still poorly understood, although there have been a few studies that have proposed explanations (Nitsche et al., 2003; Nitsche et al., 2005). During stimulation, changes in excitability are assumed to be accomplished by changing the membrane potential within neurons (for review, see Stagg & Nitsche, 2011). Evidence for this comes from studies in which chemically blocking calcium or sodium channels abolishes the effects of anodal tDCS on cortex during stimulation (Nitsche et al., 2003). Aftereffects of anodal tDCS may also result from synaptic modulation as a result of changes in synaptic plasticity dependent on GABAergic and glutamatergic systems (Nitsche et al., 2005). Effects of typical tDCS protocols last for about an hour after stimulation (Nitsche & Paulus, 2000). These aftereffects are longer lasting than TMS, which allows only short-term increases in cortical excitability (Ziemann et al., 2008).

The most typical design of a tDCS experiment is to apply stimulation to one group of participants and to do between-groups comparisons with a sham-stimulation control group, although some studies have used multi-day within-subjects designs. A major advantage of tDCS in comparison to TMS is that this sham-stimulation condition is thought to be a more convincing placebo, given that participants actually receive a short 30 second burst of stimulation during the beginning and end of the stimulation period that evokes similar sensations to the actual stimulation procedure. Given the amount of time required for tDCS to change the targeted region's neural membrane potential, this 30 second burst is not enough to produce any physiological effects. Because of its advantages over TMS and the ease of its administration, the experiments in this dissertation use tDCS to explore questions about episodic retrieval.

Brain Stimulation and Memory

The series of experiments in this dissertation is the strongest test to date of the hypothesis that dlPFC plays a causal role in episodic retrieval. For the tDCS literature, I searched the Web of Science database with the terms: "tDCS' or 'transcranial direct current stimulation" AND "episodic memory" or "recognition memory." This search provided 51 results. Of these results, I selected and carefully reviewed the 25 empirical papers that applied either anodal or cathodal tDCS to either dlPFC or temporoparietal regions in healthy populations for an episodic memory task. I repeated the same process for the TMS literature, with the terms: "TMS' or 'transcranial magnetic stimulation" AND "episodic memory" or "recognition memory." This search provided 127 results. Of these results, I carefully selected and reviewed the 27 empirical papers that applied TMS to either dlPFC or temporoparietal regions in healthy populations for episodic memory tasks as well. I also conducted broader searches for both tDCS and TMS with just the term "memory" and read through the titles and abstracts to make sure I did not miss any potentially relevant study (the tDCS search had 555 results and the TMS search had 1,728 results). Of all of the published studies, the experiments in this dissertation are by far the most comprehensive test of the hypothesis that dlPFC has a causal role in episodic retrieval in terms of number replication attempts and number of participants. We collected data from over 350 different participants across various stimulation conditions in several different memory tasks. In Experiment 4, we had 40 participants in each stimulation condition alone, which is more than any of the studies found in this literature search. Thus, any conclusions that can be drawn from this series of experiments should have a profound impact on our understanding of how tDCS to dlPFC before retrieval influences memory.

The majority of studies using tDCS or TMS to modulate prefrontal cortex function have demonstrated how stimulation of this region affects performance on working memory tasks (e.g., Jones, Gozenman, & Berryhill, 2015; Martin et al., 2014; Coffman, Clark, & Parasuraman, 2014; Richmond, Wolk, Chein, & Olson, 2014; Mulquiney, Hoy, Daskalakis, & Fitzgerald, 2011; Zachle et al., 2011; Fregni et al., 2005; see Kuo & Nitsche, 2012 for review). Although these experiments do not test longer-term episodic memory, many aspects of working memory are critical for both encoding and retrieval processes. For example, on a source memory task, a person needs to hold the targeted memory cue in mind while simultaneously evaluating whether they have enough evidence to determine if they saw the item in a particular source.

A recent meta-analysis by Brunoni and Vanderhasselt (2014) stated that while TMS reliably affects working memory capacity (see also Luber & Lisanby, 2014), tDCS only leads to improvements in reaction time, but not in accuracy or error rates. Kuo and Nitsche (2012; 2015) also emphasize that tDCS enhances reaction time, but instead proclaim that some tDCS studies *do* find significant benefits for working memory accuracy, although the results are inconsistent. Berryhill, Peterson, Jones, and Stephens (2014) suggest tDCS effects on dlPFC in working memory tasks are subtle and that these inconsistencies in the literature are a result of underpowered studies (particularly in between-subjects designs), heterogeneous populations that add unnecessary variance, and insufficiently challenging cognitive tasks. These factors were taken into account when designing the studies presented in this dissertation.

In contrast to the vast number of studies using brain stimulation over prefrontal cortex to modulate working memory, there have been relatively few studies about episodic memory (see Manenti, Cotelli, Robertson, & Miniussi, 2012 for a partial review). Of the studies that have been published, it is critical to separate them into two types: those in which stimulation is applied

before or during encoding, and those in which stimulation is applied after encoding and before (or during) retrieval. Assuming the memory test is given within a short time frame after encoding, brain stimulation will affect both encoding *and* retrieval processes, as well as any short-term reconsolidation effects between study and test. This means that, while encoding stimulation studies are useful for making general claims about the importance of brain regions for memory, the conclusions about how they affect specific processes are limited. Still, such studies still can provide evidence of the importance of various brain regions (such as prefrontal, temporal, and parietal cortex) during basic episodic learning tasks.

Effects of Temporoparietal Stimulation on Episodic Memory

Stimulation during encoding

Although the main purpose of this dissertation is to study the effects of stimulation on prefrontal cortex, is it nevertheless important to also consider the effects of stimulation on temporoparietal regions, which are strongly connected to prefrontal cortex. Studies applying TMS or tDCS before or during encoding to temporoparietal cortex generally find significant differences between stimulation and control. Wang and Voss (2015) and Wang and colleagues (2015) used fMRI to localize the region of lateral prefrontal cortex that is most strongly connected to hippocampus in each of their participants during a memory task. Stimulating this region via repetitive TMS (rTMS) led to increased memory accuracy for arbitrary face-word pairings. Pisoni and colleagues (2015) found that anodal tDCS of parietal or temporal regions led to increases in d' on a word recognition task. Jacobson, Goren, Lavidor, and Levy (2012) applied anodal tDCS to left superior parietal cortex with the cathode located on the same region in the right hemisphere, and also found increased d' on similar word recognition task. Jones, Gözenman, and Berryhill (2014) similarly found that anodal stimulation of left posterior parietal

cortex led to increased recall scores on a list of words presented auditorily. Boggio and colleagues (2009) found evidence that bilateral stimulation of the anterior temporal lobe (left anode/right cathode) or unilateral stimulation of left anterior temporal lobe with a large, highly distributed cathode over right temporal lobe led to decreases in false alarms to critical lures on a DRM memory task (Roediger & McDermott, 1995). The only study that did not find memory differences as a result of temporoparietal stimulation was by Rossi and colleagues (2006), which applied TMS to left or right parietal cortex but did not find changes in memory for visual scenes. This literature as a whole suggests applying TMS or tDCS to temporal and parietal cortex regions before encoding leads to changes in memory accuracy.

Stimulation before or during retrieval

The majority of studies applying TMS or tDCS to temporoparietal regions at retrieval also find significant effects on memory accuracy. Pergolizzi and Chua (2015) applied bilateral tDCS (anode on left posterior parietal cortex; cathode on right posterior parietal cortex) before retrieval and found decreases in false alarms to critical lures on a DRM task. Chi, Fregni, and Snyder (2010) applied bilateral tDCS (both combinations of left/right, anodal/cathodal) stimulation over temporal cortex for lists of pictures of shapes all connected by a theme and measured performance on the task, within subjects, before and after stimulation conditions. Memory discrimination scores for the visual stimuli increased most after right anodal/left cathodal stimulation. Sestieri and colleagues (2013) applied rTMS to the superior parietal lobe or angular gyrus (or appropriate sham control conditions) before retrieval, however, and found that while stimulation of angular gyrus led to lower recognition than superior temporal lobe, neither condition differed significantly from sham. In sum, the few studies apply TMS or tDCS over temporoparietal regions before retrieval found significant improvements in memory accuracy.

Effects of Prefrontal Stimulation on Episodic Memory

Stimulation during encoding

In general, TMS of prefrontal regions during encoding leads to increases in episodic memory accuracy. Turriziani and colleagues (2010) found that rTMS over left dlPFC during an encoding task in which participants made "pleasantness" judgments led to decreased memory accuracy for verbal stimuli, whereas rTMS over right dlPFC led to disruptions in memory for non-verbal stimuli. Blumenfeld, Lee, and D'Esposito (2014) applied rTMS to either left dlPFC or left ventrolateral PFC (vlPFC) during encoding for a memory task in which participants were asked to judge whether words were concrete or abstract. Stimulation of vlPFC led to decreases in item recognition (via increased false alarms to lures), although stimulation of dlPFC did not disrupt memory performance. Kohler, Paus, Buckner, and Milner (2004) used rTMS to stimulate left inferior prefrontal cortex during a task in which participants studied a list of nouns. Participants who received left inferior prefrontal cortex stimulation remembered *more* words than those in two other stimulation control condition. Hawco, Berlim, and Lepage (2004) found evidence of individual differences in the effects of rTMS over left dlPFC, such that highly strategic participants (measured by a pre-experiment questionnaire) in a word pair memory task had their memory performance decrease following stimulation, whereas low strategy participants had their memory performance increase following stimulation. Disruptions in memory for visuospatial stimuli occurred as a result of rTMS to right dlPFC in younger adults in research by Rossi and colleagues (2004; 2006), and in memory for faces (Tuzziziani et al., 2008). rTMS of right dlPFC led to decreased performance on a memory task for words and shapes in which no judgment was made during encoding (Blanchet, Gagnon, & Schneider, 2010). In sum, the TMS literature largely suggests left dlPFC stimulation during encoding influences memory

performance for word stimuli, whereas right dlPFC disruption affects the encoding of nonverbal stimuli.

Several studies have applied tDCS to dlPFC during encoding, but the results have been considerably less consistent than those of the TMS literature. Javadi and colleagues found, for both standard long-duration (Javadi & Walsh, 2012) and short duration, trial-based bursts (Javadi, Cheng, & Walsh, 2012), anodal tDCS led to increases in accuracy and speed on a multiple choice recognition verbal memory task, whereas cathodal tDCS led to decreases in accuracy. Sandrini and colleagues (2016) applied anodal tDCS over left dlPFC to a group of older adults while they were studying a list of 20 words. They found that subjects who received anodal tDCS remembered more words from the list 48 hours and 30 days later. Lu, Wang, Chen, and Xue (2015) demonstrated that anodal stimulation of left dlPFC led to improved accuracy on a recognition memory task for verbal visual forms (Korean characters). They also demonstrated that anodal stimulation enhanced spatiotemporal neural pattern similarity (measured by EEG) in the contralateral hemisphere across repetitions of the same item, which, in turn, predicted better memory performance. Other studies, however, have found no benefits of anodal stimulation or even decreases in accuracy as a result of tDCS stimulation on dlPFC before retrieval. Nikolin and colleagues (2015) applied high definition tDCS (HD-tDCS) to left dlPFC immediately before a verbal learning task in which they were asked to immediately recall a list of nouns presented to them, but found no differences between stimulation and sham in accuracy on this task. Furthermore, they found no performance differences in recognition tests given the next day. Elmer and colleagues (2009) applied either anodal or cathodal stimulation to left or right dlPFC, but did not find any differences between stimulation and sham in recall of auditory stimuli. Penolazzi and colleagues (2010) applied bilateral stimulation to dlPFC in which one hemisphere

was anodal and the other cathodal (both combinations were conditions) before having participants study pictures of varying emotional arousal/valence. Although there was an interaction between valence and stimulation condition on memory performance, no difference was found for overall memory of the pictures between stimulation and sham conditions. Finally, Zwissler and colleagues (2014) demonstrated that anodal stimulation over left dlPFC led to increases in false memory for pictures of scenes that were similar in gist to some of those studied during encoding, whereas cathodal stimulation of this region decreased false alarms to these items. In other words, recollection accuracy was *decreased* from anodal stimulation in this study. In sum, while the majority of TMS studies have shown that stimulation to temporoparietal regions before encoding have relatively consistent effects on memory accuracy, research using tDCS to apply stimulation to these regions has led to inconsistent results across studies.

Stimulation before or during retrieval

In contrast to effects of applying TMS to prefrontal cortex before encoding, there is little evidence that stimulating prefrontal cortex before or during retrieval influences accuracy, although some studies have suggested that there may be effects on response latencies. Gagnon, Blanchet, Grondin, and Schneider (2010) showed that paired-pulse TMS over right dlPFC, but not left dlPFC, led to increases in discrimination for both verbal and nonverbal materials. However, they were later unable to replicate this accuracy difference (Gagnon, Schneider, Grondin, & Blanchet, 2011), although they did find decreased reaction times as a result of right dlPFC stimulation compared to left dlPFC. Manenti and colleagues (2010) applied rTMS to left and right dlPFC and found no differences in memory for abstract words between stimulation and sham, although stimulation increased reaction time for both right and left dlPFC stimulation compared to sham. Rylas and colleagues (2016) applied theta-burst stimulation over frontopolar

cortex or dlPFC just before an associative object recognition task. They found no evidence of accuracy increases as a result of this stimulation, but did find that frontopolar stimulation led to more accurate judgments of learning for subsequently forgotten items.

Some studies, however, have found evidence suggesting individual differences in TMS effects on prefrontal cortex that may account for the lack of accuracy differences found in heterogeneous populations. Manenti and colleagues (2010) divided individuals into "high strategy" and "low strategy" users in an association task consisting of both celebrity and unknown name-face associations. They applied rTMS to left and right dlPFC and found that right dlPFC stimulation led to accuracy decreases for strategy users compared to sham, whereas left dlPFC stimulation led to accuracy decreases in non-strategy users. Similarly, Manenti, Cotelli, and Miniussi (2011) applied rTMS to left and right dlPFC in older adults during encoding or retrieval of word pairs. Groups were separated into low-performing and high-performing groups. For low-performing groups, left dlPFC stimulation during encoding led to decreases in memory for word pairs compared to sham, but increases in performance when applied during retrieval. For high-performing groups, stimulation made no difference during encoding or retrieval. It is important to consider, however, that these analyses were likely conducted post-hoc, and taken together, the overall consensus of the TMS retrieval literature is that increasing or disrupting prefrontal function with TMS does not reliably affect memory retrieval.

The few studies applying tDCS over dlPFC regions after encoding but before retrieval have also not found convincing evidence of its role in retrieval. The studies described here were all published after the current work was initiated. Manenti and colleagues (2013) applied anodal stimulation to either left or right dlPFC during an old/new recognition memory task in both

younger and older adults, but did not find any differences between stimulation and sham in accuracy for either group. Jones, Gozenman, and Berryhill (2014) applied stimulation over left dlPFC before retrieval, but did not find increases in the number of words recalled on an auditorily presented list. Morgan, Davis, and Bracewell (2014) found no effects of bilateral stimulation of dlPFC on recognition accuracy for emotional picture stimuli. One study did find evidence that cathodal stimulation over dlPFC might affect retrieval accuracy. Penolazzi and colleagues (2014) applied anodal, cathodal, or sham tDCS over right dlPFC to participants during the retrieval practice phase in a standard recall retrieval-induced forgetting paradigm using word stimuli. Anodal stimulation did not affect accuracy for standard recall targets, although cathodal stimulation eliminated the retrieval-induced forgetting effect. These results suggest inhibiting right dlPFC reduces the retrieval monitoring processes that lead to the retrieval-induced forgetting effect.

A few recent studies have suggested, however, that anodal tDCS over dlPFC might lead to changes in memory consolidation that, in turn, lead to later memory improvements. Two studies applied anodal stimulation over left dlPFC in younger adults (Javadi & Cheng, 2013) or older adults (Sandrini et al., 2014) during the reactivation period of a long-term memory consolidation paradigm and found increases in accuracy for reactivated items compared to sham. Marshall and colleagues (2004) applied bilateral anodal stimulation to left and right dlPFC during slow wave sleep and found evidence of improvements on a word-pair memory task in stimulated participants, but did not find evidence of anodal tDCS effects in a similar experiment in which stimulation was applied during wake. It may thus be the case that in the short term, tDCS over dlPFC has little effect when applied before retrieval, but instead has its effects on the reconsolidation processes that occur long after the study phase is over.

As a whole, TMS and tDCS studies do not provide convincing evidence that stimulation of dlPFC before retrieval leads to reliable changes in memory accuracy. This does not mean that dlPFC does not play a role in retrieval monitoring; none of the studies included in this review used memory tasks that specifically targeted recollection accuracy for different kinds of source information. We might expect, based on the neuroimaging literature, that if dlPFC contributes to episodic memory retrieval, the effects would be strongest when people recollect specific kinds of source information and must engage in cognitively controlled retrieval monitoring processes. Across the four experiments presented in this dissertation, we applied tDCS to prefrontal cortex just before retrieval to address the role of this region in recollection and the associated retrieval monitoring processes.

Project Motivation

The major goal of this project was to provide evidence of a causal role of dlPFC during episodic memory retrieval. To do so, we applied stimulation to dlPFC after encoding but before retrieval of the criterial recollection task. This allowed us to explore the contributions of dlPFC regions to episodic recollection of specific kinds of source information. Given that previous neuroimaging work suggests a correlation between dlPFC activity and retrieval monitoring using the same criterial recollection task (Gallo, Kensinger, & Schacter, 2006; Gallo, McDonough, & Scimeca, 2010; McDonough, Wong, & Gallo, 2012), this design should be well suited to make causal claims about the specific contributions of dlPFC in the context of this previous research.

No research to date has used tDCS to study recollection of specific source information and the associated retrieval monitoring processes, let alone tried to resolve the laterality debate about how each hemisphere contributes to retrieval. If dlPFC subserves these retrieval monitoring processes, then tDCS should increase the quality of memories, enabling people to

more accurately recollect the specific details of studied items (Koriat, Goldsmith, & Pansky, 2000) and be less likely to falsely recall incorrect details. Furthermore, by applying tDCS to either left or right dlPFC, we also aimed to provide a breakdown as to how the two hemispheres differentially contribute to retrieval.

Understanding how the prefrontal cortices are involved in retrieval is critical for our understanding of memory more generally. Consider eyewitness testimony, a situation in which memory retrieval accuracy is at a premium. Understanding the importance of prefrontal retrieval strategies could, for example, allow us to discount the testimonies of participants who are not carefully monitoring each face in an eyewitness identification task. Perhaps this research could even lead to the use of advanced imaging techniques, such as predictive ERP or pattern-based fMRI analysis, to develop a precise neural "signature" of accurate eyewitness identification (or at least to rule out individuals who are guessing or unsure). Furthermore, it has been hypothesized that as people age, they become more reliant on prefrontal retrieval strategies (Davis et al., 2008). Understanding the involvement of prefrontal cortex in these strategies is critical to developing better mnemonics to maximize accuracy in aging individuals.

The goal of Experiment 1 was to provide an extension of fMRI findings in Gallo, McDonough, and Scimeca (2010) using tDCS to 1) determine if tDCS can lead to improved retrieval on a criterial recollection task and 2) to explore prefrontal laterality for detail retrieval. Experiments 2 and 3 were designed to replicate the results of Experiment 1 while manipulating task difficulty. Experiment 4 was designed to both directly replicate Experiment 1 and to extend the findings to an eyewitness identification task. We also conducted a pilot experiment that attempted to replicate the findings of Experiment 1 in a two-alternative forced choice (2AFC)

paradigm, as well as to explore metacognitive confidence changes as a result of stimulation (see Appendix A).

EXPERIMENT 1

The goal of Experiment 1 was to provide an extension of Gallo, McDonough, and Scimeca (2010) using tDCS as way to test the correlational claims from their fMRI project. All participants went through the same version of the criterial recollection task used in this fMRI study. Rather than going into an fMRI scanner during retrieval, however, participants were administered 20 minutes of tDCS before retrieval in one of four conditions/locations, according to the 10-20 EEG coordinate system: stimulation to left dlPFC (Anode: F3, Cathode: right supraorbital region), stimulation to right dlPFC (Anode: F4, Cathode: left supraorbital region), control stimulation to left parietal cortex (Anode: P5 [just lateral of P3], Cathode: right supraorbital region), or control sham stimulation (identical to either left or right sham stimulation set-up, but with no actual stimulation provided).

Methods

Recruitment and Inclusion Criteria

Participants were initially given a safety survey before participating in the study to confirm that it was safe to perform brain stimulation. This survey confirmed that participants were not pregnant, had no history of seizures or strokes, and were not currently taking any psychoactive drugs or medications at the time of participating. Additionally, all participants recruited for this experiment were right-handed. Although we asked for participants' handedness in the pre-screening process, we confirmed that participants were right-handed by checking to make sure that participants' scores were 40 or higher on the Edinburgh Handedness Inventory

(EHI; Oldfield, 1971). All participants included in the final sample of this experiment fit these criteria.

Participants

A total of 96 (48 male) participants met initial inclusion criteria and participated for either class credit or $40 compensation (24 in left stimulation, 24 in right stimulation, 24 in sham, and 24 in parietal control stimulation). We chose to use this number of subjects in each condition because it is similar to the number of participants used in comparable tDCS studies (e.g. Pergolizzi & Chua, 2015; Manenti et al., 2013), as well as the sample size used in our previous between-subjects work with older adults (e.g. McDonough, Wong, & Gallo, 2012; Gallo, Foster, & Johnson, 2009). All participants were between the ages of 18-30 and were students at the University of Chicago. In addition to these 96, 7 participants were unable to complete the study due to inability to properly set up tDCS because of hair extensions (3 participants), minor discomfort (burning sensation) during stimulation (2 participants), or falling asleep during the protocol (2 participants).

tDCS

tDCS was delivered by a battery-driven stimulator (Soterix Medical, New York, NY, United States). Two conductive-rubber electrodes were placed inside 5 cm x 7 cm (35 cm^2) sponges that had been dampened with a 0.45% saline solution. Before stimulation began, all participants received approximately 30 seconds of the stimulation ramping up to and down from 1 mA of as a preview of the stimulation in order to condition the skin for further stimulation and to allow for adjustments if necessary. In the active stimulation conditions, 2 mA of current (0.057 mA/cm^2 current density) was administered to participants for 20 minutes. During the sham stimulation conditions, current was ramped up to and down from 2 mA over the course of

about 30 seconds at the beginning and end of the 20-minute stimulation period. This was intended to simulate a convincing stimulation sensation for participants, but to not provide stimulation for a long enough period of time to have any lasting effects.

Participants were randomly assigned to one of four conditions: stimulation of left dlPFC (n = 24), stimulation of right dlPFC (n = 24), or sham stimulation of either left or right dlPFC (n = 24 total – 12 left, 12 right), or control stimulation of the left parietal lobe (n = 24). This last condition was included to confirm that any effects of frontal tDCS found were regionally specific and not due to brain stimulation in general. In the left dlPFC stimulation and sham conditions, the active anodal electrode was positioned at electrode position F3 according to the 10-20 electroencephalographic-system (Jasper, 1958), with the cathodal electrode placed on the right supraorbital region. Stimulation over the supraorbital region is physiologically inert. In the right dlPFC stimulation and sham conditions, the active anodal electrode was instead positioned at electrode position F4, with the cathodal electrode placed on the left supraorbital region. These electrode locations have been used in other memory studies using dlPFC stimulation (e.g., Javadi & Walsh, 2012; Boggio et al., 2009). In the left parietal stimulation control condition, the active anodal electrode was placed just laterally of position P3 (to correspond to position P5 in higher channel electroencephalographic caps), an area shown to be involved in memory (but not monitoring) in several other studies using ERP (e.g., Vilberg & Rugg, 2009), TMS (e.g., Wang et al., 2014), and tDCS (e.g., Pergolizzi & Chua, 2015; Manenti et al., 2013). The cathode was placed on the right supraorbital region.

Materials and Procedure

Diagnostic and disqualifying monitoring ability were measured using a criterial recollection task procedure identical to that of Gallo, McDonough, and Scimeca (2010) (see

Figure 4 for a visual depiction of the general procedure for Experiments 1-4). Before beginning the criterial recollection task, participants completed a practice version of the task, using separate stimuli, to ensure they understood the instructions (approximately 5 minutes). Each study trial began with a black word (500 ms), followed (after a 100 ms interstimulus interval) by the same word in larger red letters (1200 ms) or the corresponding picture (1200 ms), with a 150 ms interstimulus interval. Participants studied 240 unique red words and pictures for the upcoming tests (90 red words, 90 pictures, 60 items presented as both red words and pictures, nonconsecutively). Each red word and picture was presented twice, nonconsecutively. Participants were instructed to press a button to make a corresponding semantic judgment for each red word ("Can this item be made in a factory?") or to make a corresponding perceptual judgment for each picture ("Is this a highly detailed image of the object?"). To avoid carryover effects of these orienting tasks, participants studied alternating blocks of red words and pictures (3 red word blocks and 3 picture blocks of 100 stimuli each), with stimuli randomized within each block and block order counterbalanced across participants. Each block was separated by 8 seconds of fixation.

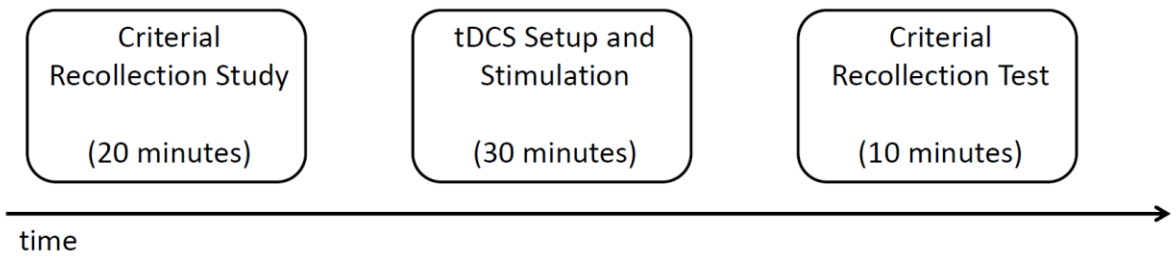
After completing the study phase but before retrieval, participants were given either real or sham stimulation. In this experiment, stimulation occurred in a different room than the criterial recollection task. Between set-up time and stimulation duration, the entire tDCS procedure took approximately 40 minutes (with some variation), factoring in the extra time it took to walk between the two rooms. The purpose of applying stimulation after encoding but before retrieval was to enhance frontally-dependent retrieval monitoring processes on the subsequent tasks.

After stimulation, the criterial recollection test phase began. Verbal labels (black font on white background) were used as retrieval cues, along with a test prompt to keep participants on task ("red word?" for the Red Word test, "picture?" for the Picture Test, and "red word only?" for the Exclusion test). The tests were divided into three "runs" to parallel the procedure used in Gallo, McDonough, and Scimeca (2010). Each run was subdivided into three test blocks corresponding to each of the three types of test (Red Word, Picture, or Exclusion) and the entire procedure took approximately 10 minutes to complete. Test block order varied across runs and was counterbalanced across participants. On the Red Word test, participants pressed "yes" if they remembered studying a corresponding red word (i.e., red word and both items) and "no" if not, regardless of whether or not they remembered a corresponding picture (i.e., picture and new items). On the Picture test, participants pressed "yes" if they remembered studying a corresponding picture (i.e., picture and both item) and "no" if not, regardless of whether they remembered the corresponding red word (i.e., red word and new items). This nonexclusion logic was strongly emphasized in the test phase instructions so that participants would focus on whether they could recollect the item in the to-be-remembered format (a diagnostic monitoring process). On the Exclusion test, participants were instructed to press "yes" if they remembered the item as a red word and "no" if not, but it was emphasized that both items would not be included on this test so that red words and pictures were mutually exclusive. Thus, if they recollected a picture on the Exclusion Test, they could be sure the item was not associated with a red word at study and could immediately reject it (a disqualifying monitoring process). During each test block, participants saw 10 words corresponding to each type of studied item (red word, picture, both, or new), with the exception that "both" items were replaced with 10 additional new items during the Exclusion Test blocks. In total, there were 30 items of each critical type (red

words, pictures, or new) on each test, with 30 filler items also included to manipulate exclusion demands (30 both items on the Red Word and Picture tests, 30 additional new items on the Exclusion test). The procedure was self-paced with a 150 ms interval between trials.

Additionally, upon arriving and before leaving the study, participants were administered a copy of the Positive and Negative Affect Scale-X (PANAS-X) (Watson & Clark, 1994). This scale was included as a control for two of its subscales: attentiveness and fatigue. More specifically, if tDCS stimulation increases levels of arousal (evidenced by increases in attentiveness or decreases in fatigue before and after stimulation), then it could be the case that these factors, as opposed to the stimulation itself, would be responsible for differences in memory performance.

Figure 1. General Procedure of Experiments 1-4



Results

Recollection Accuracy

For the primary analysis we calculated an accuracy score for each of the three recollection tests in each stimulation condition (i.e., left, right, sham, and parietal, see Figure 2). This accuracy score (hits minus false alarms) reflected participants' ability to discriminate between target words that had been associated with the criterial format at study (e.g., picture items on the Picture Test) and lure words that had been associated with the nontarget format at

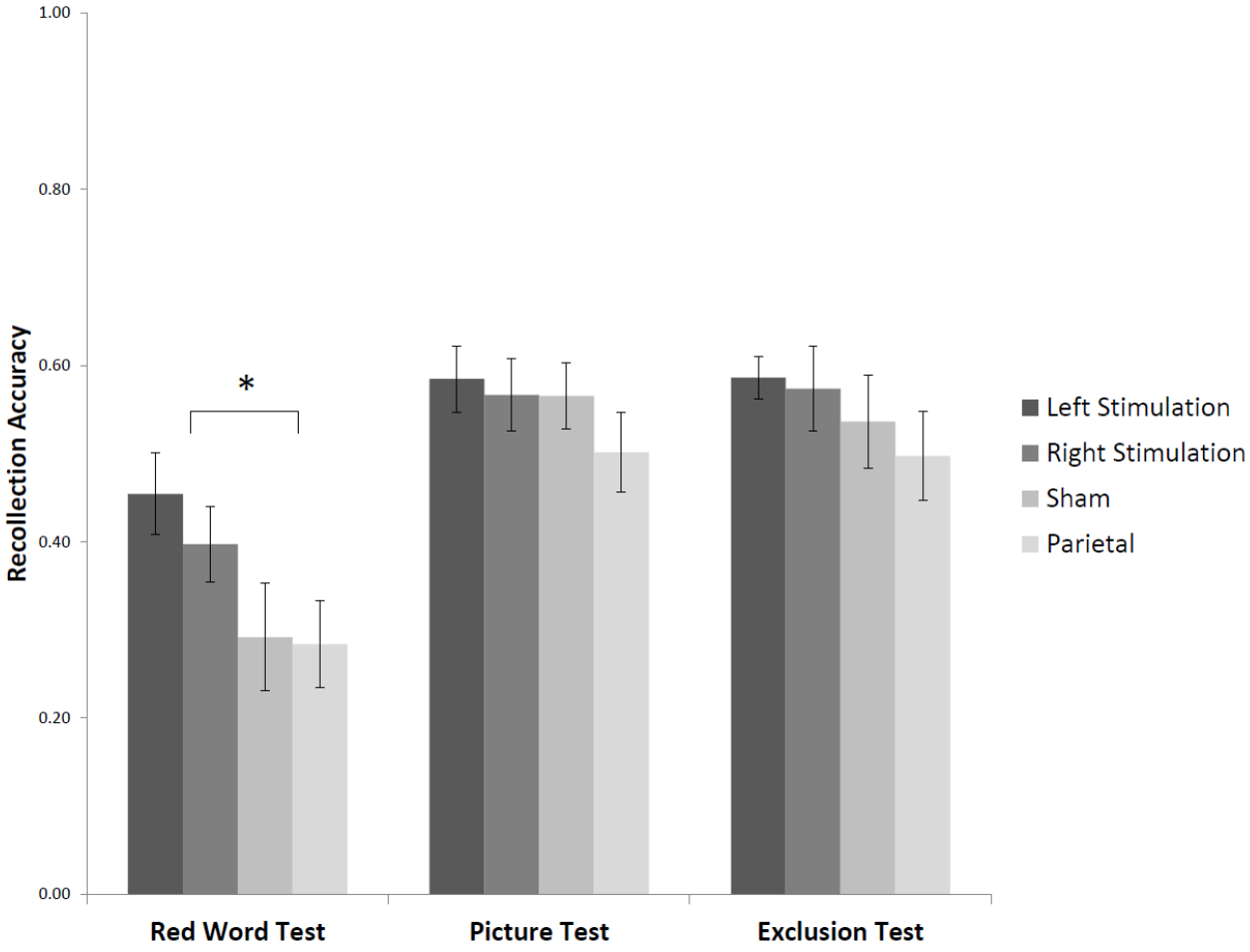
study (e.g., red font items on the Picture Test), while disregarding responses to items that had been associated with both formats as well as nonstudied items. As discussed by Gallo (2013), the targets and the lures used in this accuracy measure should have been similar in familiarity from the study phase, so participants needed to use recollection to differentiate them at test. Moreover, because this recollection accuracy score reflects participants' ability to discriminate between the same kinds of red font and picture items on each of the three recollection tests, differences in this accuracy score across tests can be attributed to differences in retrieval demands. Consistent with these assumptions, overall accuracy (collapsing across dlPFC stimulation conditions) was lower on the Red Word Test than the Picture Test, $t(71) = 6.90$, $p < .001$, $d = .85$, demonstrating the expected advantage of distinctive picture recollections on performance. Accuracy also was lower on the Red Word Test than the Exclusion Test, $t(71) = 7.41$, $p < .001$, $d = .76$, demonstrating the advantage of using picture recollections in an exclusion process. There was no accuracy difference between the Picture Test and the Exclusion Test ($t < 1$), as each of these tests benefited from distinctive picture recollections.

To justify looking breaking down the data by individual test, we wanted to determine if there was a significant difference in stimulation effect size across the red word, picture, and exclusion tests. We conducted an omnibus 3 (test: red word, picture, exclusion) x 2 (stimulation: active, sham) mixed ANOVA and found a significant effect of test, $F(2, 140) = 41.45$, $p < .01$, $\eta 2p = .37$, no effect of stimulation, $F(1, 70) = 1.80$, $p = .18$, $\eta 2p = .03$, and critically, a significant interaction between test and stimulation, $F(2, 140) = 3.00$, $p = .05$, $\eta 2p = .04$. As predicted, we found different stimulation effect sizes across the three test types, such that the largest stimulation effects were present on the Red Word Test.

We next ran a 2 (stimulation: active, sham) x 2 (hemisphere: left, right) ANOVA on each of the three recollection tests, considering only the conditions with electrodes placed on dlPFC sites. This analysis revealed that recollection accuracy on the Red Word Test was improved by tDCS, as there was a main effect of stimulation (active > sham), $F(1, 68) = 4.65$, $p = .04$, $\eta 2p = .06$, but no main effect of hemisphere (left, right), and no interaction (p's > .25). Follow-up t-tests revealed that there was a main effect of left stimulation compared to sham, $t(46) = 2.14$, $p = .04$, $d = .62$, but no effect of right stimulation, $t(46) = 1.42$, $p = .16$, $d = .41$. In contrast, stimulation had no significant effects on the Picture Test or the Exclusion Test (all p's > .19). Stimulation of dlPFC increased recollection accuracy on the Red Word Test more than the other two tests, which showed no benefit from stimulation.

As an additional test of regional specificity, we compared recollection accuracy scores in the parietal stimulation condition to the sham stimulation condition. Unlike the analysis of dlPFC stimulation, this analysis of parietal stimulation revealed no significant differences in accuracy scores on any of the three tests relative to the sham condition (all p's > .29). Moreover, a direct comparison between the parietal stimulation condition and the dlPFC stimulation conditions (collapsing hemispheres) revealed that recollection accuracy on the Red Word Test was significantly greater with dlPFC stimulation than parietal stimulation, $t(70) = 2.51$, $p = .01$, $d = .62$, with no differences in accuracy on the other two tests (both p's > .14). Taken together, these analyses demonstrate that parietal stimulation did not yield the same performance benefit as did dlPFC stimulation.

Figure 2. dlPFC Stimulation and Recollection Accuracy in Experiment 1



Note. The * represents significance at the α = .05 level, two-tailed.

Hits and False Alarms

In addition to recollection accuracy scores, we also analyzed performance separately for targets and lures in each of the brain stimulation conditions (see Table 1). Unlike recollection accuracy scores, which theoretically represent the ability to discriminate between targets and lures independent from overall response bias effects that may vary across brain stimulation conditions, a separate analysis of hits and false alarms might be affected by unintended differences in response bias across brain stimulation conditions. However, a separate analysis of

targets and lures within the standard (no stimulation) condition can provide important information about distinguishing between memory search and postretrieval monitoring effects.

Before turning to stimulation effects on targets and lures, we first describe the behavioral results from the sham (no stimulation) condition, which replicated four key results from Gallo, McDonough, and Scimeca (2010). First, on each of the three tests, participants responded "yes" more often to targets than to lures (all p's < .01), demonstrating that they used recollection to differentially respond to test items that had been associated with red font or pictures at study. Second, on each of the three tests, participants made significantly more false alarms to lures that were studied in the incorrect format (e.g., picture items on the Red Word Test and Exclusion Test; Red Words on the Picture Test) compared to nonstudied items, demonstrating an effect of stimulus familiarity on false recollection errors (all p's < .05). Third, false alarms to both studied lures and nonstudied lures were greater on the Red Word Test than on the Picture Test (both p's < .01), demonstrating a distinctiveness effect on false recollection errors and that retrieval monitoring demands were greater on the Red Word Test than the Picture Test. Fourth, participants made significantly fewer false alarms to studied lures on the Exclusion Test compared to the Red Word Test, $t(23) = -5.33$, $p < .01$, $d = -.95$, but there was no reduction in false alarms to nonstudied lures, $t(23) = 1.44$, $p = .16$, $d = .26$. This selective reduction in false alarms on the Exclusion Test demonstrates that participants had used picture recollections to reduce recollection errors via an exclusion process, again implicating increased retrieval monitoring demands on the Red Word Test.

To investigate brain stimulation effects on targets and lures separately, we compared hits and false alarms across the dlPFC stimulation conditions. For the Red Word Test, a 2 x 2 ANOVA revealed a main effect of stimulation (active, sham), $F(1, 68) = 10.10$, $p < .01$, $\eta 2p =$

.13, no effect of hemisphere, $F(1, 68) = 0.08$, $p = .78$, $\eta 2p < .01$, and no interaction, $F(1, 68) = 0.09$, $p = .08$, $\eta 2p < .01$. Follow-up t-tests revealed that left dlPFC stimulation significantly boosted hits to red font items compared to sham stimulation, $t(46) = 2.46$, $p = .02$ $d = .71$, with no effect on false alarms to picture items, $t(46) < 1$. Similarly, right dlPFC stimulation significantly boosted hits to red font items compared to sham stimulation, $t(46) = 2.77$, $p = .01$, $d = .80$, with no effect on false alarms to picture items, $t(46) < 1$.

In contrast, there were no significant differences in hits or false alarm rates between stimulation and sham conditions for the other two tests (all p's > .22), nor did parietal stimulation significantly affect these hit or false alarm rates relative to sham (all p's > .30). These analyses indicate that the benefits of dlPFC stimulation on recollection accuracy on the Red Word Test was primarily driven by a significant increase in hit rates to targets, as opposed to a reduction in false alarms to lures.

Response Latencies

In addition to performance metrics, we also analyzed response latencies separately for hits and correct rejections. First looking at the difference between tests themselves, regardless of stimulation, participants were faster on the Picture Test than the Red Word Test at both identifying targets, $t(72) = 5.22$, $p < .01$, $d = .39$, and correctly rejecting noncriterial lures, $t(72) = 4.42$, $p < .01$, $d = .47$. They were also faster on the Red Word Test than the Exclusion Test at identifying targets, $t(72) = 5.51$, $p < .01$, $d = .40$, but there was no difference in speed at rejecting picture lures, $p > .25$. This pattern is generally consistent with previous literature (e.g., Gallo, McDonough, & Scimeca, 2010) suggesting that participants are faster on the Picture Test because picture stimuli are more distinctive and require less processing.

Next, we looked at hits to targets on the three tests. A 2 x 2 ANOVA revealed no effect of stimulation (active, sham) on latency for red word hits on the Red Word Test, $F(1, 68) = 2.42$, $p = .12$, $\eta 2p = .03$, no effect of hemisphere (left, right), $F(1, 68) = 2.65$, $p = .11$, $\eta 2p = .04$, and no interaction, $F < 1$. There was a trending effect of stimulation on the Picture Test, $F(1, 68) = 3.17$, $p = .08$, $\eta 2p = .05$ such that participants were slower with stimulation compared to sham, but no effect of hemisphere or an interaction (both F's < 1). There were no effects of stimulation, hemisphere, or an interaction on the Exclusion Test (all F's < 1).

Then, we looked at correct rejections to noncriterial lures. A 2 x 2 ANOVA revealed a main effect of stimulation (active, sham) on latency for correct rejections of picture stimuli on the Red Word Test, $F(1, 68) = 4.44$, $p = .04$, $\eta 2p = .06$, as participants were *slower* in the active condition than the sham condition. This suggests tDCS may have motivated participants to spend more time searching and monitoring during each trial. There was no effect of hemisphere (left, right) or an interaction on the Red Word Test, (both F's < 1). On the Picture Test, there was no main effect of stimulation, $F(1, 68) = 2.01$, $p = .16$, $\eta 2p = .03$ and no effect of hemisphere or an interaction (both F's < 1). On the Exclusion Test, there was no effect of stimulation ($F < 1$), no effect of hemisphere, $F(1, 68) = 1.44$, $p = .24$, $\eta 2p = .02$, and no interaction between the two ($F < 1$).

Table 1: Mean Proportion of "Yes" Responses on Each Recollection Test as a Function of Brain Stimulation Condition and Response Latency (ms) in Experiment 1

	Probability of "yes"				Latencies for Hits/CRs			
	Left	Right	Parietal	Sham	Left	Right	Parietal	Sham
Red Word Test								
Both	0.74 (.03)	0.72 (.03)	0.69 (.04)	0.69 (.03)	1477	1712	1370	1537
Red Words	0.77 (.03)	0.79 (.03)	0.67 (.05)	0.65 (.04)	1505	1655	1410	1377
Pictures	0.31 (.04)	0.40 (.04)	0.39 (.05)	0.35 (.04)	1886	2041	1679	1564
New	0.08 (.01)	0.16 (.02)	0.18 (.03)	0.10 (.02)	1472	1640	1429	1352
Picture Test								
Both	0.84 (.02)	0.82 (.02)	0.73 (.03)	0.79 (.03)	1338	1352	1194	1135
Pictures	0.73 (.03)	0.74 (.03)	0.65 (.04)	0.70 (.03)	1363	1384	1308	1202
Red Words	0.14 (.02)	0.17 (.02)	0.15 (.02)	0.14 (.02)	1520	1605	1434	1396
New	0.04 (.01)	0.05 (.01)	0.06 (.02)	0.05 (.01)	1459	1406	1288	1196
Exclusion Test								
Red Words	0.74 (.03)	0.77 (.03)	0.68 (.04)	0.72 (.03)	1870	2212	1939	1906
Pictures	0.16 (.02)	0.19 (.03)	0.18 (.03)	0.18 (.03)	1766	1940	1590	1726
New	0.11 (.02)	0.25 (.04)	0.25 (.05)	0.13 (.02)	1709	1824	1622	1496

Note. Standard errors of each mean are in parentheses. CRs = correct rejections. Red words were targets on the Red Word Test and Exclusion Test, and lures on the Picture Test. Pictures were targets on the Picture Test, but lures on the Red Word Test and Exclusion Test. Latencies are only for correct responses ("yes" for targets, "no" for lures).

Arousal Analyses

In order to determine if the difference in memory discrimination was related to increases in arousal, we conducted a follow-up analysis using the attentiveness and fatigue subscales of the PANAS-X (Watson & Clark, 1994) to determine if there were any differences between pre- and

post-stimulation levels of these two factors selective to either group. If there is either increased attentiveness or decreased fatigue as a result of stimulation, but not sham, then it could be the case that the effects were not due to regionally specific stimulation, but instead an overall change in arousal levels.

A 2 (stimulation: left v. sham) x 2 (attentiveness: pre v. post) mixed ANOVA did *not* reveal a main effect of pre v. post attentiveness, $F(1, 45) = 2.24$, $p = .14$, $\eta 2p = .05$, suggesting pre and post attentiveness levels were similar before and after stimulation, regardless of stimulation or sham condition. There was also no main effect of stimulation, $F < 1$, such that the two groups did not differ in overall attentiveness levels. Critically, there was also no interaction between stimulation condition and pre v. post attentiveness levels, $F < 1$. These analyses were also done with fatigue levels as the dependent variable. There was no main effect of fatigue, $F < 1$, meaning that across both groups, participants were not more tired after the experiment than when they came in. There was, surprisingly, a group difference in overall fatigue levels, $F(1, 45) = 4.85$, $p = .03$, $\eta 2p = .10$, such that the sham group had higher overall levels of fatigue than the left stimulation group (more on this below). Critically, there was again no interaction between stimulation and pre v. post fatigue levels. The lack of interactions in these analyses suggests neither stimulation nor sham conditions selectively changed arousal levels after the stimulation procedure, and thus supports the specificity of stimulating left dlPFC for finding memory changes.

Because there was an unexpected group difference in overall fatigue levels between the left stimulation and sham groups, we conducted a follow-up analysis to determine if this difference was responsible for the group memory differences. To do so, we repeated the analysis for red word discrimination above and controlled for pre-stimulation fatigue levels as a co-

variate. While this is not a perfect measure, given the way the experiment was designed, we cannot look at the difference between fatigue levels pre- and post-stimulation because post-stimulation fatigue was assessed after completion of the memory test, which in itself likely influenced this measure (arousal is more carefully controlled for in Experiment 4). The 2 (stimulation: active, sham) x 2 (hemisphere: left, right) ANOVA revealed a significant effect of stimulation, $F(1, 64) = 4.21$, $p = .04$, $\eta 2p = .06$, no effect of hemisphere, $F(1, 64) = 0.02$, $p = .88$, $\eta 2p < .01$, and no interaction, $F(1, 64) = 1.15$, $p = .29$, $\eta 2p = .02$. Thus, given these findings and that the parietal stimulation did not lead to memory improvements, we can be confident that the group differences in memory performance were due to the stimulation procedure (and not attentiveness or fatigue levels).

Discussion

Experiment 1 demonstrated that electrically stimulating prefrontal cortex improves retrieval, allowing people to more accurately recollect the details of their experiences. These tDCS effects were behaviorally selective, as stimulation only boosted recollection accuracy when people were focused on recollecting less distinctive information, when retrieval monitoring demands were high (i.e., the Red Word Test compared to the Picture and Exclusion Tests). These tDCS effects also were regionally selective, as stimulation of left parietal cortex did not benefit recollection accuracy, even though both regions have been associated with memory retrieval. Finally, these results were not due to increases in arousal levels as a result of stimulation, as there was no interaction between self-reported arousal levels and stimulation condition. Taken together, these results suggest that dlPFC may play a causal role in retrieval.

The finding that dlPFC stimulation benefitted the Red Word Test more than the Picture Test is consistent with several neuroimaging studies using this task, which have repeatedly

suggested dlPFC is more active on the Red Word Test than the Picture Test (e.g., Gallo et al., 2006; 2010). We also found that tDCS stimulation to left or right dlPFC increased recollection accuracy on the Red Word Test. Prior fMRI work has focused on right dlPFC activity during the Red Word Test (Gallo et al., 2010), but the Red Word Test often activates both left and right dlPFC relative to the Picture Test. More generally, neuroimaging and brain damage studies have yielded mixed evidence on the laterality of retrieval processes (e.g., Nolde, Johnson, & Raye, 1998; Cabeza, Locantore, & Anderson, 2003). In the current study we were able to differentially target each of the two hemispheres with tDCS, and the findings indicate that both left and right dlPFC are causally involved in retrieval. Because we found that stimulation effects in both hemispheres were related to increases in hits and not false alarms, this suggests that stimulation of either hemisphere improved memory search, rather than postretrieval monitoring. The relationship between increased activity in dlPFC and memory search contradicts the predictions of the *production-monitoring* hypothesis. However, these results also conflict with the left-systematic/right-heuristic account, which would have predicted to see larger effects of right dlPFC stimulation on the word task than left stimulation. To confirm our findings, we explored this laterality question further in Experiment 4.

In contrast to the stimulation effects we observed on the Red Word Test, we found that dlPFC stimulation did not significantly benefit performance on the Exclusion Test. The use of distinctive picture recollections in the exclusion process reduces retrieval demands, and stimulating dlPFC was not expected to benefit performance on the Exclusion Test as much as the Red Word Test. However, the fMRI study of Gallo, McDonough, and Scimeca (2010) associated the Exclusion Task with dlPFC activity, suggesting that this test might have benefitted from tDCS to some extent. It may be that the dlPFC is causally involved in the exclusion process, but

that using picture recollections in this exclusion process is not too demanding of dlPFC resources, so that dlPFC stimulation does not yield additional performance benefits. Alternatively, it may be that the dlPFC activity observed with fMRI was not causally linked to the use of an exclusion process, but instead reflected some correlated aspect of processing. Either of these interpretations is viable. Because of this uncertainty, we decided to move away from the exclusion task in Experiments 2-4.

The primary analysis of brain stimulation effects focused on recollection accuracy scores, because this measure is designed to control for familiarity effects as well as differences in response bias that may affect responding to both targets and lures across conditions. However, by breaking down recollection accuracy scores into hits and false alarms, we can gain insight into whether stimulation was affecting the memory search process (hits) or the postretrieval monitoring process (false alarms). In this experiment, the effects of dlPFC stimulation on the Red Word Test were primarily driven by increased hit rates to targets as opposed to reduced false alarm rates to lures. This pattern suggests dlPFC stimulation increased the effectiveness of the memory search process of retrieval (which by definition is associated with targets but not lures). In contrast, this pattern is inconsistent with the idea that dlPFC stimulation increased the effectiveness of postretrieval monitoring. In other words, this provides support that dlPFC may be involved in early selection processes through which retrieval is constrained on the front-end during memory search (Jacoby, Kelley, & McElree, 1999), which is unsurprising given the general overlap of left PFC activity during encoding and retrieval.

Regardless of this interpretation, the results of Experiment 1 clearly suggest tDCS at retrieval can increase recollection accuracy, demonstrating a potential causal role of dlPFC in retrieval. Are the selective stimulation effects on the Red Word Test, but not the Picture Test, a

direct result of increased cognitive demands on the Red Word Test or because of item format (i.e., left dlPFC is not involved in searching and monitoring for items studied as pictures)? To get at this question of cognitive demands, we explicitly manipulated global task difficulty across Experiments 2 and 3.

EXPERIMENT 2

The goal of Experiment 2 was to replicate the results of Experiment 1 using a less demanding version of the criterial recollection task (taken from McDonough, Wong, & Gallo, 2012). This version of the criterial recollection task was used because we had originally planned to try and extend this research to elderly participants, who would have struggled with the speeded version of the task in Experiment 1. In this version, participants studied the words and pictures for a longer duration of time than in Experiment 1. Retrieving details about words should be more demanding than retrieving details about pictures, even in this easier version of the task. If stimulation of dlPFC is selectively enhancing retrieval for more demanding stimuli, then we would expect to find selective performance boosts in accuracy on the Word Test, as we did in Experiment 1. In Experiment 2, stimulation was delivered only to left dlPFC (Anode: F3, cathode: right supraorbital region), as these are where the strongest effects in Experiment 1 occurred.

Methods

Recruitment and Inclusion Criteria

The pre-screening process was identical to Experiment 1: participants were administered the same safety survey and handedness scale (EHI). In Experiment 2, we administered a different consent form than that of Experiment 1. The critical difference from Experiment 1 was that this consent form informed participants that they may or may not be in the sham condition. We

assumed that some participants would "figure out" that there was a sham condition regardless of whether we informed them about it, and that this would level the playing field across all participants. To test this assumption, we will compare the results of Experiment 2 using the new consent form to those of Experiment 1.

Participants

A total of 48 (18 male) participants were recruited to participate in this experiment and met initial inclusion criteria for either class credit or $40 compensation (24 in left stimulation, 24 in sham). Given the significant results in Experiment 1, we assumed that choosing an identical number of participants in each condition would give us sufficient power to detect another stimulation effect. In addition to these 48, 2 individuals were unable to complete the study due to minor discomfort (burning sensation) during stimulation (1 participant) or ignoring task instructions (1 participant). All participants were between the ages of 18-30 (mean age: 20.19 years) and were students at the University of Chicago.

tDCS

The tDCS parameters were identical to those of Experiment 1, although there were some notable differences in the procedure. The machine used in Experiment 2 was different than that used in Experiment 1, although both machines were made by the same company (Soterix Medical, New York, NY). Also tDCS was administered in the same room as the memory procedure, unlike in Experiment 1. Furthermore, we implemented a semi-double blind procedure in Experiment 2. The experimenter was blind as to which condition subjects were in (stimulation or sham) until just before stimulation began, after instructions were given and the machine was already set up and ready to go. This was done to minimize any experimenter bias effects that could have influenced the results of Experiment 1. This setup was used in Experiments 3 and 4

as well. Stimulation and sham procedures were only administered with left dlPFC (F3) under the anode and the right supraorbital region under the cathode.

Additionally, we also added a post-stimulation questionnaire to determine how participants' expectations about the stimulation procedure affected their performance. It was administered at the end of the experiment to assess if participants could detect whether they were in the active or sham stimulation conditions. We also asked whether they thought stimulation affected their performance.

Materials and Procedure

Diagnostic monitoring ability was measured using a criterial recollection task procedure identical to that of McDonough, Wong, and Gallo (2012). No Exclusion Test was included in Experiment 2 for two reasons: 1) we wanted to target the condition with the strongest results from Experiment 1 and 2) we wanted to simplify the task for older adults. Before beginning the criterial recollection task, participants completed a practice version of the task, using separate stimuli, to ensure they understood the instructions (approximately 5 minutes). During the study phase, each study trial began with a black word (750 ms), immediately followed (100 ms) by the same word in larger red letters (3000 ms) or the corresponding picture (3000 ms), after which participants had an additional 3000 ms to press a button to make the corresponding factory judgment for words ("Can this item be made in a factory?") or detail judgment for pictures ("Is this a highly detailed image of the object?"). There was a 50 ms interstimulus interval. Participants studied 216 unique red words and pictures for the upcoming tests (72 red words only, 72 pictures only, 72 "both" items, presented as both red words and pictures, nonconsecutively). Each red word and picture was presented once. Stimuli and study block order were counterbalanced across conditions.

After completing the study phase but before retrieval, participants were given either real or sham stimulation with the tDCS stimulator. Between set-up time and stimulation duration, the entire tDCS procedure took approximately 30 minutes (with some variation).

After stimulation, the criterial recollection test phase began. Verbal labels (black font on white background) were used as retrieval cues, along with a test prompt to keep participants on task ("red word?" for the Red Word test, "picture?" for the Picture Test). Tests were divided into three "runs." Each run was subdivided into two test blocks corresponding to each of the test types (Red Word or Picture) and took approximately 8 minutes to complete. Test block order varied across runs and was counterbalanced across participants. During each test block, participants saw 12 words corresponding to each type of studied items (red word, picture, both, or new). In total across the runs, there were 36 items of each critical types (red words, pictures, or new) on each test, with 36 filler items also included to manipulate exclusion demands ("both" items on the Red Word and Picture tests). On the Red Word test, participants pressed "yes" if they remembered studying a corresponding red word (i.e., red word and both items) and "no" if not, regardless of whether or not they remembered a corresponding picture (i.e., picture and new items). On the Picture test, participants pressed "yes" if they remembered studying a corresponding picture (i.e., picture and both item) and "no" if not, regardless of whether they remembered the corresponding red word (i.e., red word and new items). This nonexclusion logic was strongly emphasized in the test phase instructions so that participants would focus on whether they could recollect the item in the to-be-remembered format (a diagnostic monitoring process). The procedure was self-paced, with 2 seconds of fixation presented after each judgment.

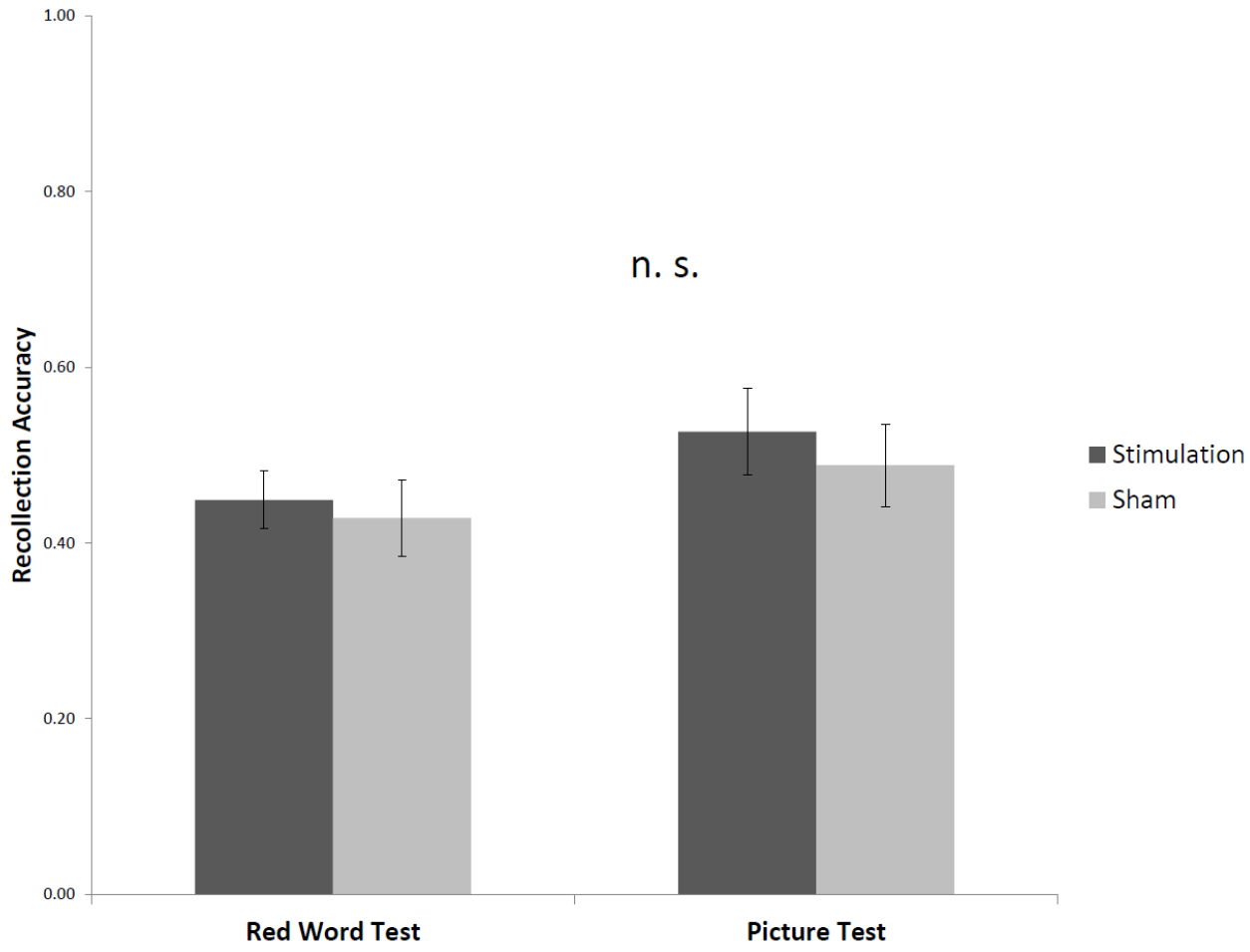
Results

Recollection Accuracy

Identical to Experiment 1, we calculated an accuracy score for both recollection tests in the stimulation and sham conditions (see Figure 3). This accuracy score (hits minus false alarms) reflected participants' ability to discriminate between target words that had been associated with the criterial format at study (e.g., picture items on the Picture Test) and lure words that had been associated with the other format at study (e.g., red font items on the Picture Test), while disregarding responses to items that had been associated with both formats as well as nonstudied items. Consistent with the distinctiveness heuristic and the results of Experiment 1, overall accuracy was lower on the Red Word Test than the Picture Test, $t(47) = 2.61$, $p = .01$, $d = .61$.

To evaluate the effects of stimulating dlPFC on retrieval, we compared recollection accuracy measures separately for both the red word and picture recollection tests. Contrary to the findings in Experiment 1, this analysis did not reveal an effect of stimulation on either the Red Word Test, $t(46) = 0.38$, $p = .71$, $d = .11$, or the Picture Test, $t(46) = 0.58$, $p = .59$, $d = .16$.

Figure 3: dlPFC Stimulation and Recollection Accuracy in Experiment 2



Note. N. s. represents non-significance at the α = .05 level, two-tailed.

Hits and False Alarms

In addition to recollection accuracy scores, we also analyzed performance separately for targets and lures (see Table 2). As in Experiment 1, the behavioral results replicated those of Gallo, McDonough, and Scimeca 2010; on both tests, participants responded "yes" more often to targets than lures (all p's < .01), made significantly more false alarms to lures that were studied in the incorrect format (e.g., picture items on the Red Word Test; Red Words on the Picture Test)

compared to nonstudied items (all p's < .05), and false alarms to studied lures and nonstudied lures were higher on the Red Word Test than the Picture Test (both p's < .05).

To investigate brain stimulation effects on targets and lures separately, we compared hits and false alarms across the dlPFC stimulation conditions. For the Red Word Test, there were no benefits of left stimulation compared to sham for hits to red word targets, $t(46) = 1.13$, $p = .26$, or in false alarms to noncriterial picture lures, $t(46) = 1.23$, $p = .22$. Similarly, there were no differences between stimulation and sham on the Picture Test for hits to picture targets, $t(46) = 0.59$, $p = .56$, or false alarms to noncriterial lures, $t(46) = 0.11$, $p = .91$. These analyses confirm that there were no notable differences found in retrieval as a result of left dlPFC stimulation.

Response Latencies

In addition to performance metrics, we also analyzed response latencies separately for hits and correct rejections. First looking at the difference between tests themselves (regardless of stimulation) participants were faster on the Picture Test than the Red Word Test at both identifying targets, $t(47) = 3.08$, $p < .01$, $d = .50$, and correctly rejecting noncriterial lures, $t(72) = 4.60$, $p < .01$, $d = .41$. This is consistent with Experiment 1.

Next, we looked at latencies for hits to targets on the Red Word and Picture memory tests. On the Red Word Test, there was no difference between stimulation and sham in response latencies for criterial hits, $t(46) = 0.57$, $p = .51$, $d = .20$. There was also no difference on the Picture Test, $t(46) = 0.84$, $p = .40$, $d = .24$. In contrast with the results of Experiment 1, we also found no differences in latencies for correct rejections of noncriterial stimuli between stimulation and sham on the Red Word Test, $t(46) = 0.22$, $p = .83$, $d = .06$, or on the Picture Test, $t(46) = 0.13$, $p = .40$, $d = .04$.

Table 2: Mean Proportion of "Yes" Responses on Each Recollection Test as a Function of Brain Stimulation Condition and Response Latency (ms) in Experiment 2

	Probability of "yes"		Latencies for Hits/CRs	
	Left	Sham	Left	Sham
Red Word Test				
Both	0.74 (.03)	0.76 (.03)	1522	1587
Red Words	0.73 (.03)	0.77 (.02)	1690	1598
Pictures	0.28 (.03)	0.34 (.04)	1783	1824
New	0.11 (.02)	0.19 (.03)	1494	1599
Picture Test				
Both	0.78 (.03)	0.76 (.03)	1364	1341
Pictures	0.66 (.04)	0.63 (.04)	1361	1477
Red Words	0.13 (.02)	0.14 (.03)	1597	1595
New	0.05 (.01)	0.06 (.02)	1453	1382

Note. Standard errors of each mean are in parentheses. CRs = correct rejections. Red words were targets on the Red Word Test and lures on the Picture Test. Pictures were targets on the Picture Test but lures on the Red Word Test. Latencies are only for correct responses ("yes" for targets, "no" for lures).

Post-experiment Questionnaire and Follow-up Analyses

According to the post-experiment questionnaire, participants in the actual stimulation condition were more likely to self-identify as having been in the stimulation condition (79%) than participants in the sham condition did (50%), $\chi^2(1, N = 48) = 4.46$, $p = .04$. This finding raises interesting methodological questions for other tDCS experiments; although we did inform participants about the nature of the sham condition (which other experiments would not typically

do), it does suggest that participants, on average, could tell the difference between stimulation and sham.

To investigate whether stimulation expectations influenced the results, we re-did the most critical analyses above looking only looking at individuals who believed they were receiving real stimulation. It should be noted that the majority of participants (79%) in this study, when asked, did not believe that stimulation (regardless of whether the actually received it) actually had any effect on their performance. Regardless, the logic here is that by only looking at individuals who believed they were being stimulated, we are eliminating potentially confounding effects of expectation (i.e., participants in sham who saw through the manipulation or overly skeptical individuals who received real stimulation). This left us with a total of 19 individuals in the stimulation condition and 12 in the sham condition. For the Red Word Test, there was no significant benefit of stimulation (mean = .47) over sham (mean = .39), $t(29) = 1.12$, $p = .27$, $d = .39$, on discrimination scores. This was also true for discrimination scores on the Picture Test (stimulation mean: .53, sham mean: .45, $t(29) = 0.83$, $p = .41$, $d = .30$. It is important to consider, however, that these analyses were drastically underpowered and that the numerical difference and medium effect sizes (particularly for Red Word discrimination) suggests the pattern *would* be significant with more participants.

Discussion

In stark contrast to the memory improvements after dlPFC stimulation found in Experiment 1, there were no significant effects of stimulation on memory detected in Experiment 2. This does not support the hypothesis that dlPFC plays a causal role in retrieval, but it is difficult to interpret null results. It may have been the case that performance on this task was high enough such that increasing prefrontal activity does not provide any further benefits

(Berryhill, Peterson, Jones, & Stephens, 2014). There also could have been larger within-group variability in our subject groups like the time of day of participation, as there was some evidence for tDCS effects on memory in the participants tested in the morning (we discuss this point in the context of this experiment and Experiment 3 in the "Time of Day Analyses" section). Alternatively, it could simply be the case that left dlPFC does not actually influence retrieval, and the finding in the Experiment 1 was a result of Type I error.

Despite the lack of memory effects, we did find differences in the post-stimulation questionnaire between the two groups. Participants receiving stimulation were more likely to believe they were receiving real stimulation than were participants in the sham group. This casts doubt on the assumption that the sham stimulation condition in a tDCS study is a perfect placebo. Because of this, we excluded participants who thought they were in sham in both groups, as this expectation may have influenced their motivation to perform well in the experiment. Although not significant due to a lack of power, this expectation-filtered analysis suggested medium effects of stimulation consistent with those of Experiment 1.

There were a number of differences between the experimental procedures that could have explained our lack of a significant stimulation effect on memory. We implemented a double-blinding procedure in this experiment that was not used in Experiment 1. Also, tDCS was conducted with a different machine than the one used in Experiment 2. Furthermore, tDCS was administered in a different room than the memory encoding and retrieval procedures in Experiment 1, whereas in this experiment they were administered in the same room. We also used a different consent form in Experiment 2 that warned participants they could be in stimulation or receive placebo stimulation. Confounding factors aside, these null results indicate

left dlPFC may not be involved when retrieval is less demanding. We conducted Experiment 3 to explicitly test the importance of dlPFC in tasks with varying retrieval demands.

EXPERIMENT 3

The difference in results of Experiments 1 and 2 suggested that task demands may matter when looking for tDCS effects. We found that stimulation of dlPFC led to improvements in the most demanding memory task in Experiment 1 (the Red Word Test), but not when the task was less demanding (due to drastically increased study time during encoding) overall in Experiment 2. The goal of Experiment 3 was to more explicitly test the difficulty hypothesis. In this experiment, item format was kept constant (i.e., participants had to remember font color for *both* tests: green or red). The difference was that green words were drastically easier to remember (i.e., seen two times each and given a factory judgment) than red words (i.e., seen one time each and given no judgment).

This study had two contrasting predictions, based on the results of the previous experiments. On one hand, if dlPFC stimulation leads to improvements on the most challenging tasks, then we would expect performance on the challenging Red Word Test to have a larger stimulation benefit than on the Green Word Test. In contrast, if left dlPFC stimulation is improving early selection "recapitulation" processes, as suggested by the improvements in memory search as a result of stimulation in Experiment 1, then we would predict that dlPFC stimulation leads to a larger benefit on the "less challenging" Green Word task, which contains an explicit factory judgment during encoding that could be used for recapitulation during retrieval.

Methods

Recruitment and Inclusion Criteria

The pre-screening process was identical to Experiments 1 and 2, and participants were administered the same safety survey and handedness scale (EHI). The consent form in Experiment 3 was identical to that of Experiment 2; the reason this was kept the same was to allow direct comparison between Experiments 2 and 3 in order to determine if the less demanding difficulty of the memory task in Experiment 2 was responsible for the null effects of stimulation (in comparison to Experiment 1).

Participants

A total of 48 (32 male) participants were recruited to participate in this experiment and met initial inclusion criteria for either class credit or $40 compensation (24 in left stimulation, 24 in sham). Because we anticipated comparing the results of this experiment to that of Experiment 2, we decided to use the same number of subjects in each condition. In addition to these 48, 3 individuals were excluded because of an inability to establish a connection (1 participant), being left-handed on the EHI (1 participant), and not being a native English speaker (1 participant). All participants were between the ages of 18-30 (mean age: 20 years) and were students at the University of Chicago.

tDCS

The tDCS parameters and procedure were identical to those of Experiment 2. We also included the same post-stimulation questionnaire about whether participants could self-identify as being in the active or sham stimulation procedures as well as whether they thought stimulation affected their performance.

Materials and Procedure

Diagnostic monitoring ability was measured using a criterial recollection task procedure. In this experiment, there was no practice version of the task because we did not want to inform participants about the "factory judgments" portion of the experiment until the very last study phase, as doing so may have led them to unintentionally make factory judgments on the no-judgment red word trials. Additionally, this also meant that our study blocks were in a fixed order: participants always studied pictures and red words with no judgments followed by a final green word study phase with factory judgments.

During the picture study phase, each study trial began with a black word (700 ms), immediately followed (100 ms) by a picture representing that item (2000 ms). Participants did not make any judgments during the picture study trials. Participants studied a total of 160 unique pictures once each. Next, during the red word study phase, each study trial began with a black word (700 ms), immediately followed (100 ms) by the same word in red font (2000 ms). Participants again did not make any judgments during the red word study trials. Participants studied 80 unique red words once each. After this, during the green word trials, each trial began with a black word (700 ms), immediately followed (100 ms) by the same word in green font (up to 2000 ms), during which time participants were instructed to press a button to make a judgment as to whether the item could be made in a factory. Participants studied 80 unique green words, with each word presented twice. There was a 700 ms interstimulus interval between stimuli of all types. After completing the study phase but before retrieval, participants were given either real or sham stimulation with the tDCS stimulator. Between set-up time and stimulation duration, the entire tDCS procedure took approximately 30 minutes (with some variation).

After stimulation, the criterial recollection test phase began. Verbal labels (black font on white background) were used as retrieval cues, along with a test prompt to keep participants on task ("green word?" for the Green Word Test, "red word?" for the Red Word Test). Tests were divided into two "runs," with each run divided into two test blocks (one Green Word Test, one Red Word Test) counterbalanced in order across participants. During each Green Word Test block, participants saw 20 items studied as green words only, 20 items studied as both green words and pictures, 20 items studied only as pictures, and 20 nonstudied items (40 of each type in total across runs). During each Red Word Test block, participants instead saw 20 items studied as red words only, 20 items studied as both red words and pictures, 20 items studied only as pictures, and 20 nonstudied items (40 of each type in total across blocks). During both tests, participants were asked to identify if they had seen the test item in the given format (i.e., in green font on the Green Word Test or in red font on the Red Word Test), regardless of whether or not they remembered seeing a corresponding picture. This nonexclusion logic was strongly emphasized in the test phase instructions so that participants could focus on whether they could recollect the item in the to-be-remembered format (a diagnostic monitoring process). The procedure was self-paced, with a 2 second interstimulus interval between trials.

Results

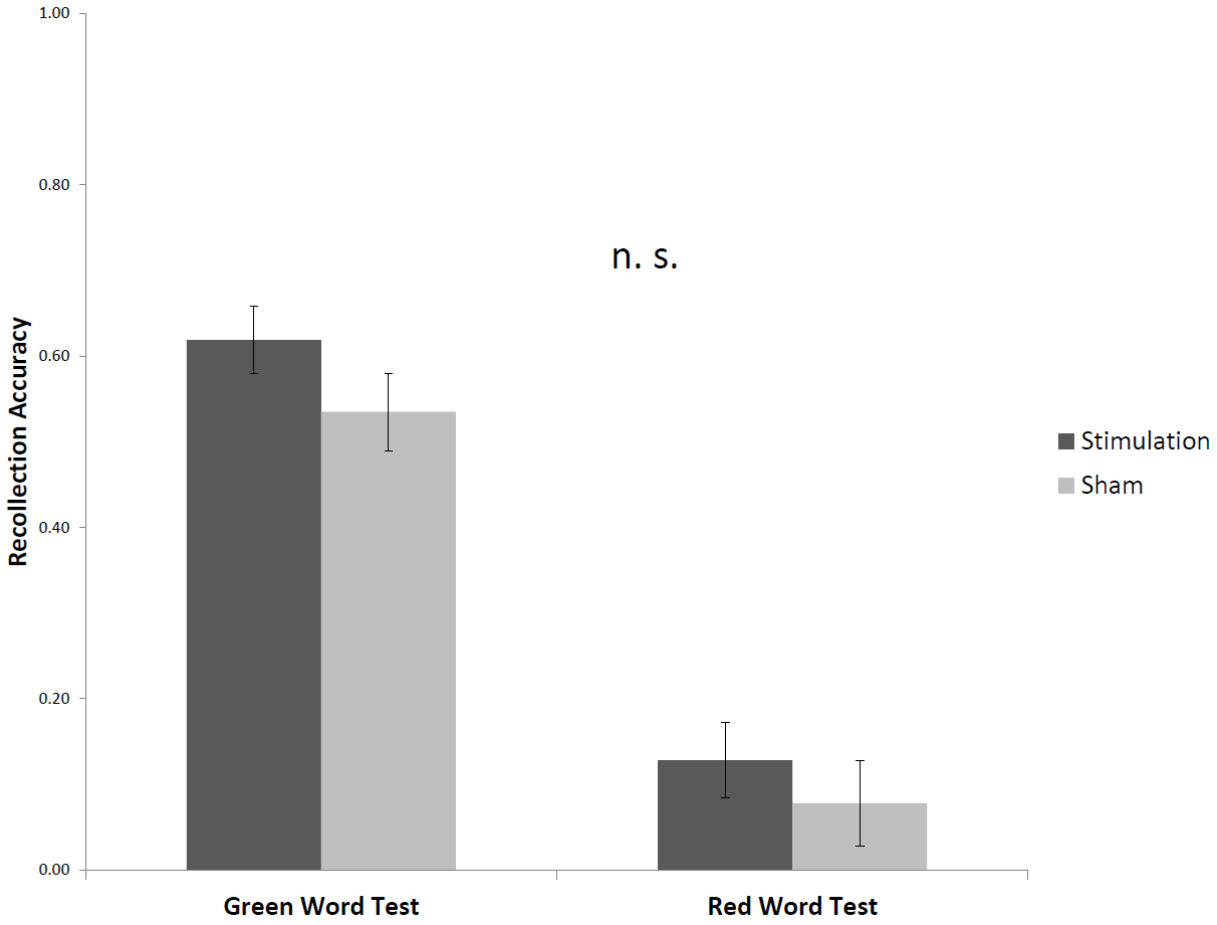
Recollection Accuracy

Identical to the analyses in Experiments 1 and 2, we calculated an accuracy score for both recollection tests in both the stimulation and sham conditions (see Figure 4), reflecting participants' ability to discriminate between target words and lure words that had been associated with pictures in the study phase. Given that green words were studied in a deeper encoding

condition than red words, it is unsurprising that overall accuracy was lower on the Red Word Test than the Green Word Test, $t(47) = 12.87$, $p < .01$, $d = 2.15$.

To evaluate the effects of stimulating dlPFC on retrieval processes, we compared recollection accuracy measures separately for both the red word and green word recollection tests. A 2 x 2 ANOVA looking at recollection accuracy as the dependent variable did not reveal a main effect of stimulation (active, sham), $F(1, 46) = 1.66$, $p = .20$, $\eta 2p = .04$. Unsurprisingly, given the difficulty manipulation, there was a main effect of test (Green Word Test, Red Word Test), $F(1, 46) = 162.75$, $p < .01$, $\eta 2p = .78$, such that green words were remembered with higher recollection accuracy than red words. Critically, there was no interaction between stimulation and test, $F(1, 46) = 0.19$, $p = .66$, $\eta 2p = .00$, suggesting that stimulation benefits to dlPFC are not affected by test difficulty.

Figure 4. dlPFC Stimulation and Recollection Accuracy in Experiment 3



Note. N. s. represents non-significance at the α = .05 level, two-tailed.

Hits and False Alarms

In addition to recollection accuracy scores, we also analyzed performance separately for targets and lures (see Table 3). Unsurprisingly, participants had higher hit rates for the Green Words than the Red Words, $t(47) = 11.75$, $p < .01$ and lower false alarm rates $t(47) = -5.09$, $p < .01$.

To investigate brain stimulation effects on targets and lures separately, we compared hits and false alarms across the dlPFC stimulation conditions. For the Green Word Test, there was no benefit of left stimulation compared to sham for hits to red word targets, $t(46) = 0.82$, $p = .42$.

There was also no benefit of stimulation on the harder Red Word Test, $t(46) = 1.50$, $p = .14$. Similarly, there were no differences between stimulation and sham on false alarm rates for either the Green Word Test, $t(46) = 1.23$, $p = .22$, or the Red Word Test, $t(46) = 0.50$, $p = .62$.

Additionally, to see if stimulation had a greater effect on one test than the other, we conducted a 2 (test type: Green Word v. Red Word) x 2 (stimulation condition: active or sham) mixed ANOVA to see if stimulation selectively boosted performance on either test. There was no interaction between test type and stimulation condition, $F < 1$. These analyses confirm that there were no notable differences found in retrieval as a result of left dlPFC stimulation in Experiment 3, and that furthermore, neither test received a selective boost.

Response Latencies

In addition to performance metrics, we also analyzed response latencies separately for hits and correct rejections. First looking at the difference between tests themselves, regardless of stimulation condition, participants were faster on the Green Word Test than the Red Word Test at both identifying targets, $t(47) = 2.16$, $p = .04$, $d = .28$, and correctly rejecting noncriterial lures, $t(72) = 9.59$, $p < .01$, $d = .88$. This is consistent with the idea that it was easier for participants to engage in retrieval on the Green Word Test than the Red Word Test.

On the Green Word Test, there was no difference between stimulation and sham in response latencies for criterial hits, $t(46) = 0.85$, $p = .40$, $d = .25$. There was also no difference on the Red Word Test, $t(46) = 0.34$, $p = .74$, $d = .10$. We also found no differences in latencies for correct rejections of noncriterial stimuli between stimulation and sham on the Green Word Test, $t(46) = 0.28$, $p = .78$, $d = .08$, or on the Red Word Test, $t(46) = 0.94$, $p = .36$, $d = .27$.

Table 3: Mean Proportion of "Yes" Responses on Each Recollection Test as a Function of Brain Stimulation Condition and Response Latency (ms) in Experiment 3

	Probability of "yes"		Latencies for Hits/CRs	
	Left	Sham	Left	Sham
Green Word Test				
Both	0.83 (.02)	0.84 (.02)	1398	1412
Green Words	0.86 (.02)	0.83 (.03)	1332	1473
Pictures	0.24 (.03)	0.30 (.04)	815	850
New	0.12 (.02)	0.07 (.02)	1444	1535
Red Word Test				
Both	0.63 (.03)	0.65 (.03)	1502	1548
Red Words	0.53 (.03)	0.46 (.04)	1534	1577
Pictures	0.40 (.03)	0.38 (.04)	1559	1789
New	0.21 (.02)	0.17 (.02)	1339	1590

Note. Standard errors of each mean are in parentheses. CRs = correct rejections. Red words were targets on the Red Word Test. Green Words were targets on the Green Word Test. Studied pictures were lures on both tests. Latencies are only for correct responses ("yes" for targets, "no" for lures).

Post-experiment Questionnaire and Follow-up Analyses

According to the post-experiment questionnaire, participants in the stimulation condition were more likely to self-identify as having been in the stimulation condition (83%) than participants in the placebo condition (42%), $\chi^2(1, N = 48) = 8.89$, $p < .01$. This replicates our finding with this questionnaire in Experiment 2 and again calls into question the validity of the tDCS sham condition in participants who are aware that they may receive sham stimulation.

To confirm that stimulation expectations did not influence the results, we re-did all of the analyses above looking only at individuals who believed they were receiving real stimulation. The majority (83%) of participants did not think that stimulation would have an impact on their performance. However, like in Experiment 2, stimulation participants were numerically, but not significantly, higher than sham participants on green word discrimination (stimulation: .62 v. sham: .54), $t(28) = 1.13$, $p = .26$, $d = .45$, and red word discrimination (stimulation: .17 v. sham: .07), $t(28) = 1.25$, $p = .22$, $d = .45$.This suggests that, with more participants who believed they were receiving stimulation, we might have found significant memory effects similar to those of Experiment 1.

Discussion

Like in Experiment 2, we were unable to replicate the effects of dlPFC stimulation on episodic retrieval. This again fails to provide evidence that dlPFC play a causal role in retrieval. However, because we used the same stimulation consent form and setup as in Experiment 2, the differences between the results of this Experiment and that of Experiment 1 are confounded.

In this study, one major caveat in interpreting these results was that the Red Word Test in this design was simply "too hard," as a number of participants were performing at floor. 29% of the stimulation participants and 42% of the sham participants had negative discrimination scores, meaning that they falsely endorsed more pictures than they successfully identified red word targets (this difference in performance was not statistically significant between groups). Participants performed significantly better on the Green Word Test (0% of stimulation participants had negative discrimination scores and only 4% of sham participants had negative discrimination scores). Because these tests had identical instructions (with the exception being to remember green words on the Green Word Test and red words on the Red Word Test), it seems

likely that participants understood the task instructions, but were simply unable to discriminate between targets and lures for the challenging red words because the task was too difficult.

It is again worth noting that, when only including participants who believed they were receiving real stimulation, stimulation participants performed numerically higher than sham participants in both tasks, with medium effect sizes. Because of these findings, we decided to go back to the original consent form that did not inform participants about the possibility of the sham condition in Experiment 4 (although we still did assess their sensations about the stimulation procedure). It is worth considering participants' expectations about the stimulation procedure in future experiments.

The null results of this study, when combined with the results of Experiment 2, do not suggest that memory task demands matter when looking for tDCS effects. Perhaps there needs to be a certain level of difficulty for dlPFC stimulation to provide a detectable "boost" in performance, but we were unable to find evidence of this across two memory tasks of very different difficulties. We conducted a final test of the difficulty hypothesis in a two-alternative forced choice pilot experiment (see Appendix A).

TIME OF DAY ANALYSES

Within tDCS studies, there is a large degree of variability within stimulation groups that can affect the efficacy of tDCS (López-Alonso et al., 2014). One suggested factor that could lead to differences in the size of stimulation effects is time of day. Although some studies have demonstrated evidence of time of day influencing the size of stimulation effects in motor cortex (Ridding & Ziemann, 2010), no studies have explored how it affects performance on cognitive tasks. This is particularly important, given the large amount of evidence suggesting that younger

adults' performance on cognitive tasks is already heavily influenced by time of day (e.g., West et al., 2002).

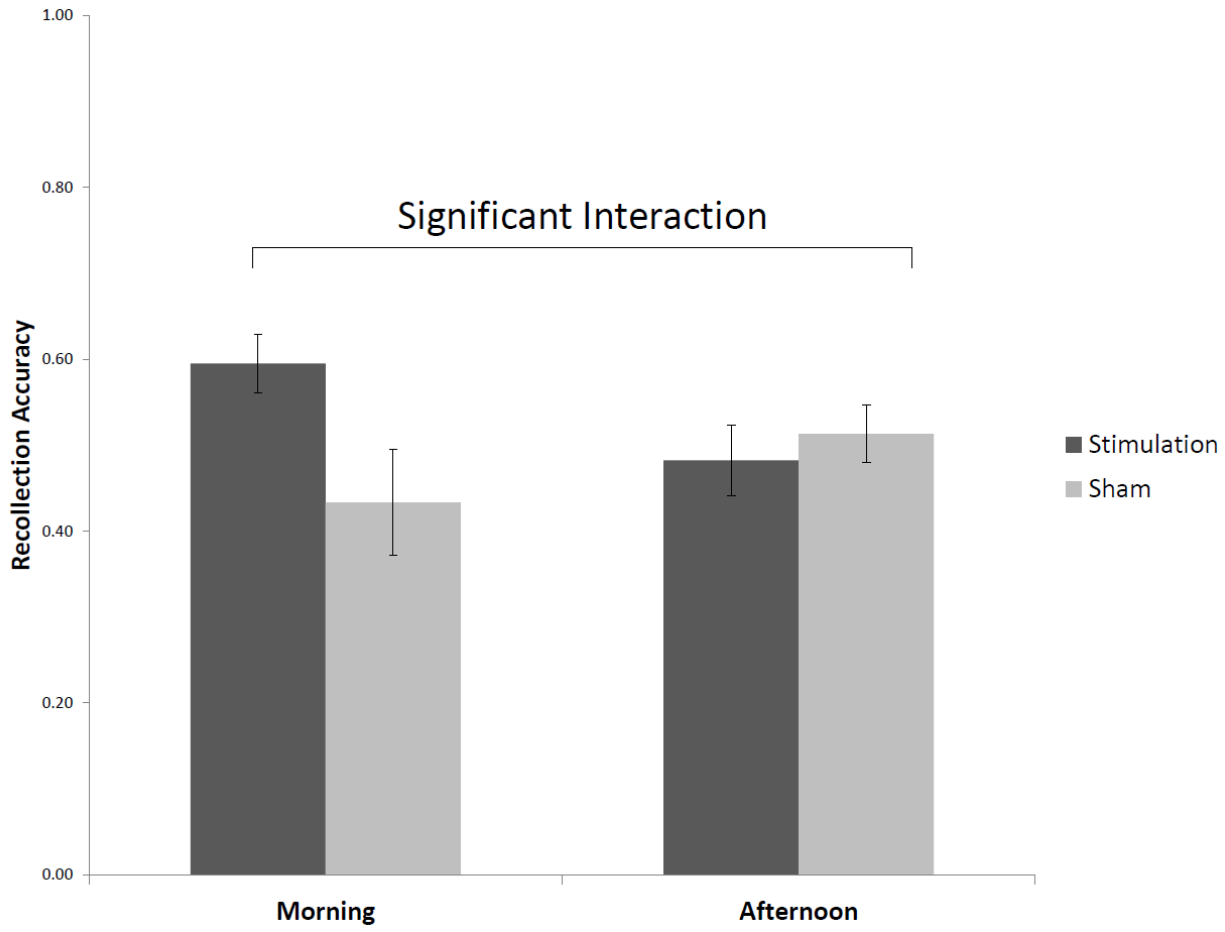
To explore time of day effects, we combined the discrimination score data from Experiment 2 on the Red Word Test and from Experiment 3 on the Green Word Test (which were identical in terms of encoding and retrieval instructions) in order to look to see if time of study participation was a confounding factor in the analyses. We did not include Experiment 1 in this analysis because of the differences in the encoding task. We split participants into two groups: morning (defined as a time slot beginning before noon) and afternoon (defined as a time slot beginning at or after noon). In Experiment 2, 46% of stimulation participants and 50% of sham participants were run in the morning. In Experiment 3, 46% of stimulation participants and 29% of sham participants were run in the morning. It is worth noting that 66% of both the stimulation and sham groups in Experiment 1 were run in the morning, so if there is an effect of time of day, it may be the case that the significant results in this Experiment were driven by this morning majority.

A 2 x 2 ANOVA showed no main effect of stimulation (active, sham), $F(1, 92) = 2.37$, $p = .13$, $\eta 2p = .03$, and no main effect of time (morning, afternoon), $F < 1$. Critically, there was a significant interaction, $F(1, 92) = 5.11$, $p = .03$, $\eta 2p = .05$, such that stimulation participants performed better than sham participants in the morning, but not in the afternoon (see Figure 5). This suggests stimulation of dlPFC had a benefit on performance overall across studies, but this effect was driven almost exclusively by our morning participants.

This is the first study to suggest time of day specificity of tDCS on a memory task. Because there are significant morning stimulation effects when combining the data from Experiments 2 and 3 in the same direction as those in Experiment 1, this may change the

interpretation of these results. It appears that these studies *do* provide evidence that left dlPFC stimulation is critical for retrieval, provided that we control for time of day. However, this finding was post-hoc; as such, we sought to replicate the finding controlling for time of day *a priori*. To account for this in Experiment 4, we only ran participants in the morning, where tDCS is expected to have its strongest effects. If nothing else, this finding demonstrates a critical need for future tDCS researchers to control for time of day in their study designs.

Figure 5. Time of Day Effects of dlPFC Stimulation in Experiments 2 and 3



EXPERIMENT 4

This research has been among the first to demonstrate that left dlPFC is causally involved in memory retrieval, at least in remembering details for word stimuli, but we were unable to

replicate these results in Experiments 2 and 3 (we also failed to replicate this effect in a pilot experiment conducted after Experiment 3, which used a different test format – see Appendix A). This may be due to variations in task instructions or demands across the experimental protocols. Although we might expect that left dlPFC is actively involved in the retrieval of memories regardless of slight variations in task instructions or demands, this has not been the case thus far, as we have only found stimulation effects on a test for details about word stimuli with a speeded study phase in Experiment 1.

A major criticism of tDCS studies has been that they are usually underpowered, between-subjects designs typically with n = 20. This leads to highly inconsistent results across the literature given the typically small effect sizes of stimulation (Tremblay et al., 2014). Given our inability to conceptually replicate the findings of Experiment 1 with slight task variations in Experiments 2 and 3, a major goal of Experiment 4 was to directly replicate the stimulation findings of Experiment 1 by using identical encoding and retrieval instructions in a sufficiently powered design.

We conducted an *a priori* power analysis using the effect sizes in word discrimination scores between the left stimulation and sham groups of Experiment 1. Given the mean difference of 0.16 between left stimulation and sham and a pooled standard deviation of 0.35, we would need to have a total of n = 40 subjects in each group to have 80% power to detect a stimulation effect – almost double the number of subjects included in Experiment 1. This number is more than any of the studies found in this literature. Furthermore, although we did not find a significant effect of right stimulation in Experiment 1 in comparison to sham, we did find a medium effect size (d = .42). If it is the case that Experiment 1 was underpowered, then it is important to include this condition in the design for Experiment 4 as well.

Thus, in Experiment 4, we included 3 stimulation conditions with n = 40 in each condition: left, right, and sham (including both left and right sham, n = 20 each). Furthermore, given the post-hoc finding in Experiments 2 and 3 that participants received a larger benefit of tDCS in the morning, we only ran participants in morning sessions (defined as between 8 AM and 12 PM). If we cannot replicate the findings of Experiment 1 under these conditions, then we need to question the reliability of those significant results.

A secondary goal of Experiment 4 was to explore the effects of tDCS on memory for faces. In our daily lives, we need to be able to remember stimuli more complicated than the color of words. Faces are among the most prominent of these stimuli to remember, given their importance for social interaction. Eyewitness testimony is a particularly challenging process through which individuals are tested on their ability to remember the face of a criminal they may have only seen for a split second at a crime scene. Failures in eyewitness testimony accuracy can lead to false convictions. Thus, the purpose of using tDCS in an eyewitness identification paradigm is twofold: first, we can provide causal evidence that dlPFC contributes to face memory, and second, we may provide evidence that tDCS itself can be a useful tool for enhancing eyewitness accuracy in a police lineup.

There have been a few empirical studies that explored the neural signatures of eyewitness memory for face stimuli (see Werner, Künnel, & Markowitsch, 2013, for review). In a somewhat similar paradigm to our own, Smirni and colleagues (2015) had participants study a series of 25 faces before applying anodal or cathodal tDCS over left or right dlPFC. They found that cathodal stimulation over right dlPFC led to improvements in memory for faces on a three-alternative forced choice test of similar faces, similar to a simultaneous eyewitness identification paradigm. They did not find any differences in their anodal stimulation condition. Lefebvre and colleagues

(2007) had participants view videos of false crimes, such as burglaries, and later asked participants to identify the culprit in a line-up of faces. Using ERP, they found evidence of a P300 in several central and parietal electrodes in response to presentations of the culprit. Iidaka, Harada, Kawaguchi, and Sadato (2012) used fMRI to explore the neural signatures of true and false memory for faces using a modified DRM paradigm with morphed faces. In this paradigm, participants studied groups of faces matched on age and gender. During the recognition test, the critical lure was a morphed face combining features of some of the faces used in the list. They found evidence of increased amygdala activity for both true and false memories of faces, as well as a correlation between anterior cingulate cortex and the reaction time difference between targets and lure trials. Using fMRI, Uncapher and colleagues (2015) found activation in ventrolateral prefrontal cortex, inferior parietal sulcus, and hippocampus (i.e., typical memory retrieval regions) when people explicitly tried to recall studied face stimuli. Together, these findings suggest a face recognition network that includes prefrontal, emotional, medial temporal, and higher order visual regions.

Given the high stakes of such judgments and the difficulty of remembering *exactly* which face was seen among a group of highly similar faces, it is likely that prefrontal cortex is involved specifically in eyewitness identification tasks. Rapcsak and colleagues (2001) demonstrated that people with frontal lobe damage were impaired in remembering novel faces compared to controls. Elgar and Campbell (2001) suggest the existence of a medial stream (including medial temporal and medial frontal regions) for encoding faces and a lateral temporofrontal stream (such as ventrolateral prefrontal cortex) for face identification. Assuming prefrontal cortex is involved in the retrieval monitoring processes involved in face recognition, then stimulating this region with tDCS may consequently lead to improvements in face discrimination.

In Experiment 4, we included an additional study phase in which participants were presented with police-like sketches of faces. They were asked to imagine that the faces belonged to criminals and to remember them for a later memory test (similar to the paradigm used in Finley and colleagues, 2015). On the subsequent memory test, participants were given single faces presented sequentially and asked to indicate whether they had studied the face earlier. Although there is still some debate as to whether sequential or simultaneous presentation of faces leads to better memory accuracy (see Wixted & Mickes, 2015 for a recent review), we chose to use sequential presentation of faces for two reasons: 1) it is the most common method for lineup presentation in police departments across the country (Wells & Olson, 2003) and thus has the most ecological validity and 2) given the failure to find any effects of prefrontal stimulation in the 2AFC pilot experiment (see Appendix A for details about word stimuli), we wanted to keep the encoding phase item presentation as similar as possible to Experiments 1-3. This also differentiates our paradigm from the simultaneous presentation procedure used by Smirni and colleagues (2015).

Given the similarity between the paradigms of Experiment 1 and this experiment for studying and retrieving words, we predicted that if dlPFC is critical for retrieval, increasing activity in this region with tDCS would lead to improvements in memory for the quality of these recollections. We also expected to find effects of right dlPFC stimulation, given the medium effect size of this condition in Experiment 1. On the eyewitness identification procedure, we expected that dlPFC stimulation would lead to better memory performance.

Methods

Recruitment and Inclusion Criteria

The pre-screening process was identical to Experiments 1-3: participants were administered the same safety survey and handedness scale (EHI). The consent form in Experiment 4 was updated to no longer inform participants about the sham condition (akin to Experiment 1). In order to coincide with our findings in Experiments 1-3 that morning participants experienced the largest stimulation effects, participants were only run in the morning (defined as before 12 PM).

Participants

A total of 120 participants (47 male) were recruited to participate in this experiment and met the initial inclusion criteria for either class credit or $20 compensation (40 in left stimulation, 40 in right stimulation, 40 in sham). As discussed above, this sample size was chosen based on an *a priori* power analysis using the effect size of Experiment 1. All participants were between the ages of 18-30 (mean age: 19.6 years) and were students at the University of Chicago. All participants were right handed according to the EHI. All participants either spoke English as a primary language or had learned it before age 6. In addition to these 120 participants, 1 participant was excluded because of discomfort with the stimulation procedure, and 4 were turned away for not meeting the inclusion criteria to participate in the study.

tDCS

In this study, participants were randomly assigned to one of the stimulation conditions: left dlPFC stimulation, right dlPFC stimulation, left sham stimulation, and right sham stimulation (similar to Experiment 1). Like in Experiments 2 and 3, the experimenters were blind as to which condition subjects were in until just before stimulation began (after stimulation instructions were

given). Left dlPFC, right dlPFC, and sham condition set-ups were identical to those used in Experiment 1.

Materials and Procedure

During encoding, participants were given two separate study blocks: the criterial recollection study phase and the eyewitness identification study phase. The order of these study blocks was counterbalanced across participants.

Diagnostic monitoring ability was measured using a criterial recollection task procedure very similar to that of Experiment 1. In this experiment, there was no practice version of the task; instead, the first few study trials of the experiment were excluded from later analysis. Each study trial began with a black word (500 ms), immediately followed (100 ms) by the same word in larger red letters (1200 ms) or the corresponding picture (1200 ms), with a 150 ms interstimulus interval. Participants studied 90 unique red words and pictures for the upcoming tests (30 as red words only, 30 as pictures only, and 30 as both red words and pictures, nonconsecutively). Each red word and picture was presented twice, nonconsecutively, with a total of 240 study trials. Participants were instructed to make a semantic judgment for each red word ("Can this item be made in a factory?") and to make a perceptual judgment for each picture ("Is this a highly detailed image of the object?"). To avoid carryover effects of these orienting tasks, participants studied alternating blocks of red words and pictures (2 red word blocks and 2 picture blocks of 60 stimuli each), with stimuli randomized within each block and block order counterbalanced across participants.

For the eyewitness identification study phase, we used a stimulus set composed of 40 sketched faces, each with a corresponding low similarity face (7 details were changed from the base face) and a high similarity face (3 details were changed from the base face). See Figure 6

for examples of each of these categories. These stimuli were taken from Finley and colleagues (2015). In the study phase, participants studied 20 of these faces for 4 seconds each (counterbalanced across conditions), each with a 500 millisecond interstimulus interval. They were told to "imagine that the faces are pictures of known criminals who are wanted by the police," and that their memory for the faces would be tested later.

After completing the study phase but before retrieval, participants were given either real or sham stimulation. Both immediately before and immediately after stimulation, participants were asked to rate their current level of attentiveness/arousal on a scale of 0-9. Between set-up time and stimulation duration, the entire tDCS procedure took approximately 30 minutes (with some variation). The purpose of applying stimulation after encoding but before retrieval was to enhance frontally-dependent retrieval processes on the subsequent tasks.

After completing stimulation, participants were given two test blocks: the criterial recollection test and the eyewitness identification test. The order of these was yoked to the same order of the two respective study phases, and was thus counterbalanced across participants.

The criterial recollection test was similar to that of Experiment 1, although we only included the Red Word Test (and not the Picture or Exclusion Tests), as this is the only test in which we found effects throughout Experiments 1-3. Verbal labels (black font on white background) were used as retrieval cues, along with a test prompt to keep participants on task ("red word?"). There were a total of 120 test trials: 30 targets studied as both a red word and a picture, 30 red word only targets, 30 picture only lures, and 30 unstudied lures. Participants pressed "yes" if they remembered studying a corresponding red word (i.e., red word only and both items) and "no" if not (i.e., picture and new items), regardless of whether or not they

remembered a corresponding picture. The procedure was self-paced with a 150 ms interval between trials.

The eyewitness test phase was similar to the sequential presentation procedure used in Finley and colleagues (2015). Participants completed a total of 60 self-paced trials: 20 target faces, 10 low similarity lures, 10 high similarity lures, and 20 unstudied faces. Participants were instructed to indicate whether they studied the face earlier (yes/no). Furthermore, they were warned that some faces were highly similar to the ones they had already studied, and should thus be careful to reject these trials. This warning was included because pilot testing showed that participants were performing at chance on this task without this warning. At the end of the two test phases, participants were administered the post-experiment questionnaire.

Figure 6: Example of Face Stimuli used in Experiment 4

Note. Face stimuli were taken from Finley et al. (2015).

Post-experiment Questionnaire

A post-experiment questionnaire was again administered to determine the rates at which participants were able to identify that they may be in a control condition. Rather than inform

participants about the sham condition directly, we asked them if they thought "electricity was being delivered to your brain," and to explain their sensations. Additionally, in order to partially control for pre-experimental arousal levels, we also asked subjects to indicate how many hours of sleep they had the previous night, the quality of this sleep, and whether caffeine was consumed during the morning of participation.

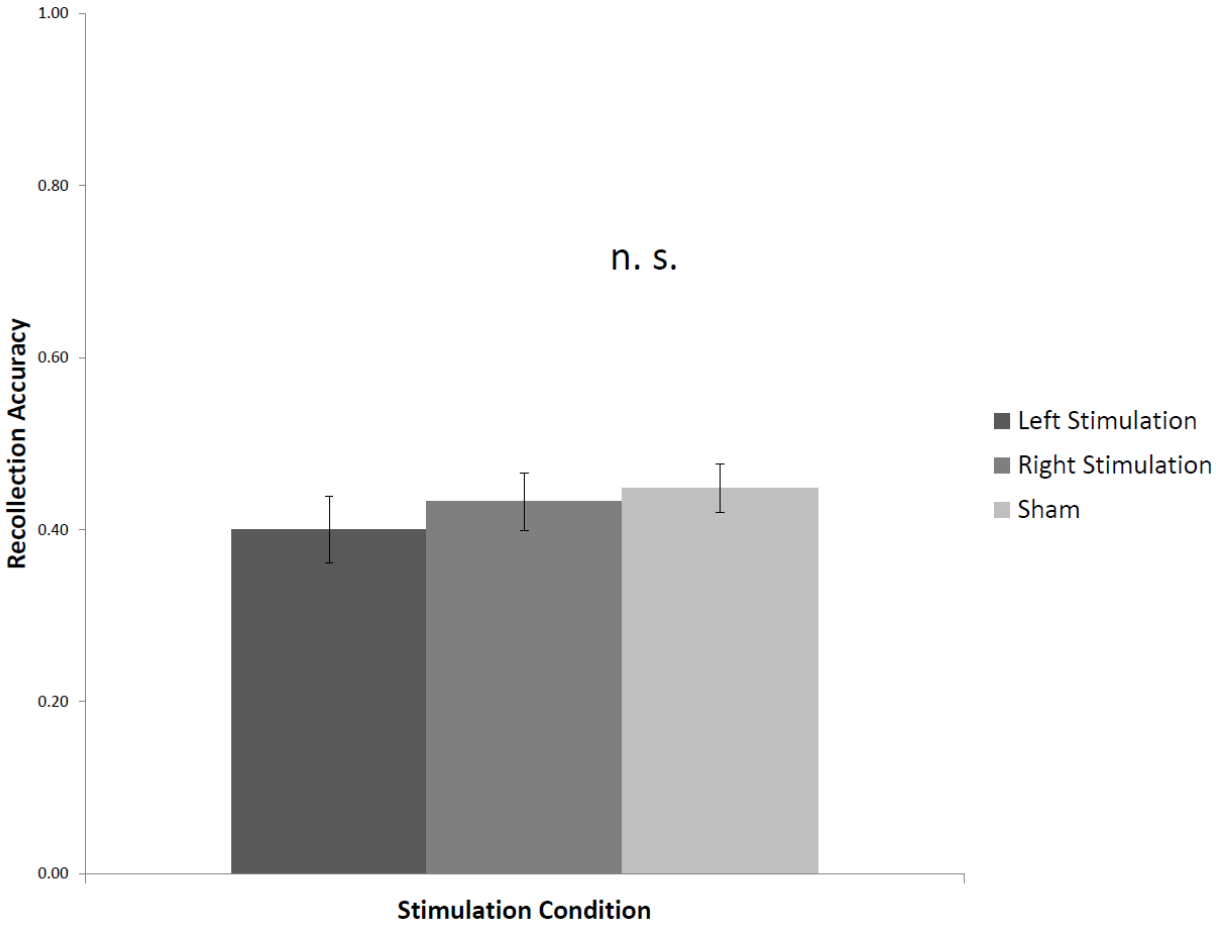
Results

Criterial Recollection Test: Recollection Accuracy

Similar to the analyses of Experiments 1, 2, and 3, we calculated an accuracy score for the red word criterial recollection test in both the stimulation and sham conditions (see Figure 7), reflecting participants' ability to discriminate between target words and lure words associated with pictures in the study phase. In addition, we included the post-stimulation arousal levels, sleep quality, sleep quantity, and the answer to a binary pre-experiment caffeine consumption question (yes/no) variable as co-variates in all of the subsequent analyses of memory performance.

To evaluate the effects of stimulating dlPFC on retrieval processes, we conducted a 2 (stimulation condition: active, sham) x 2 (hemisphere: left, right) ANOVA using recollection accuracy on the Red Word Test as the dependent variable of interest. This ANOVA did not reveal a significant effect of stimulation, $F(1, 112) = 0.58$, $p = .45$, $\eta 2p = .01$, of hemisphere, $F(1, 112) = 1.31$, $p = .25$, $\eta 2p = .01$, or an interaction between the two, $F(1, 112) = 0.11$, $p = .75$, $\eta 2p < .01$. In contrast to the results of Experiment 1, there was *no difference* between participants' memory among the stimulation conditions on the Red Word Test.

Figure 7: dlPFC Stimulation and Recollection Accuracy on Criterial Recollection Test in Experiment 4



Note. N. s. represents non-significance at the α = .05 level, two-tailed.

Criterial Recollection Test: Hits and False Alarms

In addition to recollection accuracy scores, we also analyzed performance separately for targets and lures (see Table 4). Consistent with past research using the criterial recollection task and the previous experiments, participants remembered significantly more targets than recalled false alarms, $t(119) = 21.90$, $p < .001$, demonstrating their ability to discriminate between target and lures. They made significantly more false alarms to lures that were studied as pictures than to new lures, $t(119) = 13.97$, $p < .001$.

When we conducted a 2 (stimulation condition: active, sham) x 2 (hemisphere: left, right) ANOVA, there was no difference between the stimulation conditions for hits to red word targets, $F(1, 112) = 0.01$, $p = .91$, $\eta 2p < .01$, between hemispheres, $F(1, 112) = 0.19$, $p = .67$, $\eta 2p < .01$, or an interaction, $F(1, 112) = 1.29$, $p = .26$, $\eta 2p = .01$. There was also no difference across the stimulation conditions in false alarms to picture lures, $F(1, 112) = 0.85$, $p = .36$, $\eta 2p = .01$, between hemispheres, $F(1, 112) = 0.88$, $p = .35$, $\eta 2p = .01$, or an interaction, $F(1, 112) = 0.09$, $p = .76$, $\eta 2p < .01$. These analyses confirm that there were no notable differences found in retrieval on a criterial recollection test as a result of left or right dlPFC stimulation.

Criterial Recollection Test: Response Latencies

In addition to performance metrics, we also analyzed response latencies separately for hits and correct rejections. On the Red Word Test, a 2 (stimulation condition: active, sham) x 2 (hemisphere: left, right) ANOVA revealed no differences between stimulation conditions in response latencies for criterial hits, $F(1, 112) = 0.07$, $p = .80$, $\eta 2p < .01$, hemispheres, $F(1, 112) = 0.19$, $p = .67$, $\eta 2p < .01$, or an interaction between the two, $F(1, 112) = 1.29$, $p = .26$, $\eta 2p = .01$. We also found no differences in latencies for correct rejections of noncriterial stimuli between the stimulation conditions, $F(1, 112) = 0.85$, $p = .35$, $\eta 2p = .01$, hemispheres, $F(1, 112) = 0.87$, $p = .35$, $\eta 2p = .01$ or an interaction between the two, $F(1, 112) = 0.09$, $p = .76$, $\eta 2p < .01$. In sum, it appears that neither left nor right stimulation had *any* effects on criterial recollection task performance.

Table 4: Mean Proportion of "Yes" Responses on Criterial Recollection Test as a Function of Brain Stimulation Condition and Response Latency (ms) in Experiment 4

	Probability of "yes"			Latencies for Hits/CRs		
	Left	Right	Sham	Left	Right	Sham
Red Word Test						
Both	0.75 (.03)	0.76 (.02)	0.80 (.02)	1468	1527	1439
Red Words	0.76 (.02)	0.75 (.02)	0.75 (.02)	1477	1535	1446
Pictures	0.36 (.04)	0.31 (.03)	0.30 (.03)	1853	1776	1776
New	0.11 (.02)	0.11 (.02)	0.09 (.01)	1353	1368	1381

Note. Standard errors of each mean are in parentheses. CRs = correct rejections. Latencies are only for correct responses ("yes" for targets, "no" for lures).

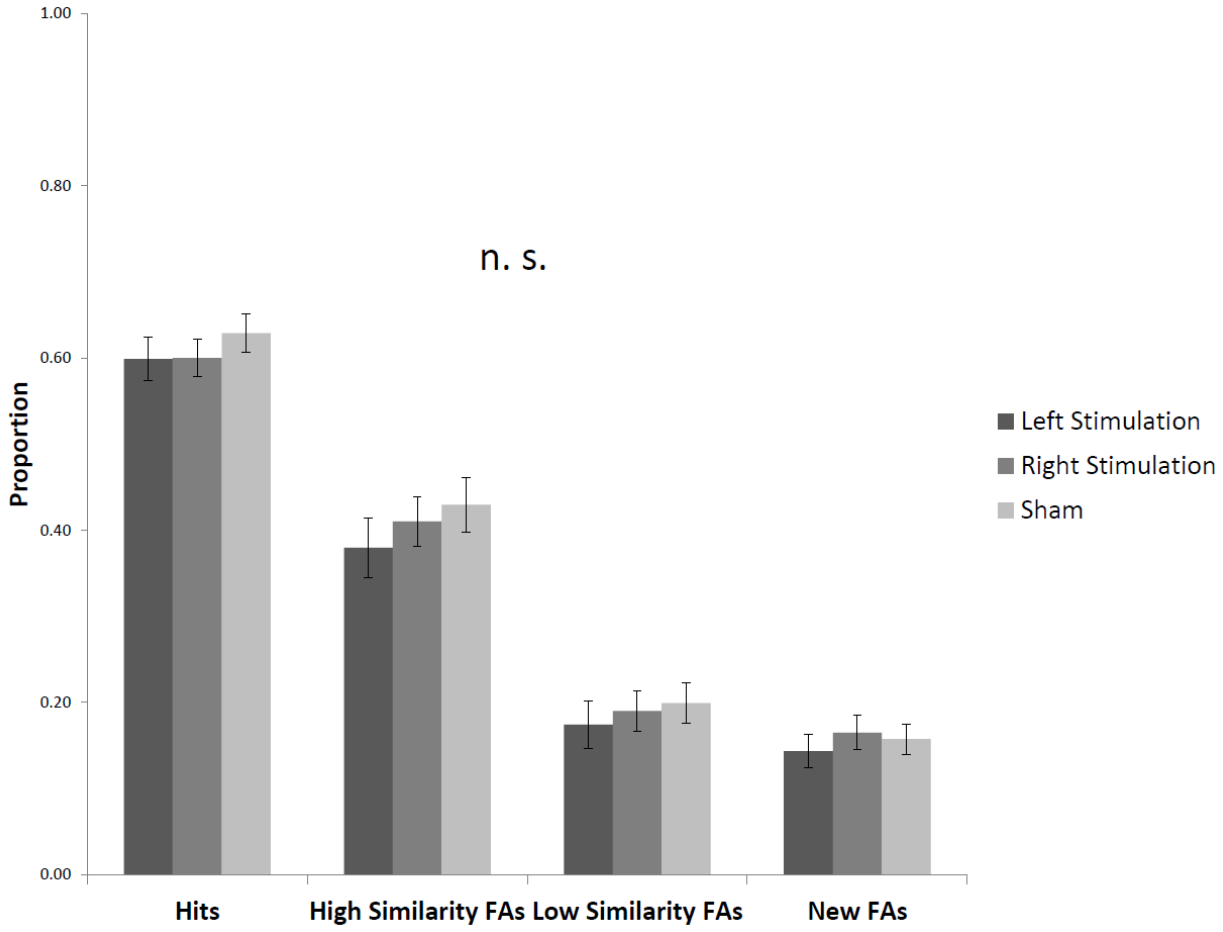
Eyewitness Identification Test: Performance

Performance during the eyewitness identification test was measured by four dependent variables: proportion of target faces remembered (hits), proportion of high similarity faces falsely remembered, proportion of low similarity faces falsely remembered, and proportion of nonstudied faces falsely remembered (see Figure 8). Given the distinct importance of false eyewitness identification in particular, we decided against combining these values into some kind of discrimination score. During the eyewitness identification test, participants were successfully able to recall target faces better than high similarity lures, $t(119) = 10.78$, $p < .001$, target faces better than low similarity lures, $t(119) = 25.24$, $p < .001$, and target faces better than nonstudied lures, $t(119) = 29.02$, $p < .001$. Furthermore, the high similarity lures were more likely than the low similarity lures to be falsely remembered, $t(119) = 10.23$, $p < .001$, demonstrating the validity of this manipulation.

We performed a 2 (stimulation condition: active, sham) x 2 (hemisphere: left, right) ANOVA across the stimulation conditions for each of these dependent variables. First looking at

hits to targets, there was no effect of stimulation, $F(1, 112) = 0.93$, $p = .34$, $\eta 2p = .01$, hemisphere, $F(1, 112) = 0.43$, $p = .51$, $\eta 2p < .01$, or an interaction, $F(1, 112) = 0.21$, $p = .65$, $\eta 2p < .01$. There was also no effect of stimulation for false alarms to high similarity face lures, $F(1, 112) = 0.92$, $p = .34$, $\eta 2p = .01$, no effect of hemisphere, $F(1, 112) = 0.49$, $p = .49$, $\eta 2p < .01$, and no interaction, $F(1, 112) < 0.01$, $p = .94$, $\eta 2p < .01$. There was no effect of stimulation for false alarms to low similarity face lures, $F(1, 112) = 0.27$, $p = .60$, $\eta 2p < .01$, no effect of hemisphere, $F(1, 112) = 1.62$, $p = .21$, $\eta 2p = .01$, and no interaction, $F(1, 112) = 0.71$, $p = .40$, $\eta 2p = .01$. Finally, there was no effect of stimulation for false alarms to new face lures, $F(1, 112) < 0.01$, $p = .98$, $\eta 2p < . 01$, no effect of hemisphere, $F(1, 112) = 2.07$, $p = .15$, $\eta 2p = .02$, and no interaction, $F(1, 112) = 0.46$, $p = .50$, $\eta 2p < .01$. In sum, stimulation of left or right dlPFC did not lead to improvements in eyewitness identification accuracy.

Figure 8: Eyewitness Memory Test Performance in Experiment 4



Note. N. s. represents non-significance at the α = .05 level, two-tailed. Hits represent correct responses to studied faces; high similarity FAs represent false alarms to high similarity lures; low similarity FAs represent false alarms to low similarity lures; new FAs represent false alarms to unstudied faces.

Eyewitness Identification Test: Latencies

We also looked at response latencies on the eyewitness identification test in order to see if stimulation had an effect on the speed with which participants made their decisions. A 2 (stimulation: active, sham) x 2 (hemisphere: left, right) ANOVA with response latencies for hits as the dependent variable revealed no differences across the stimulation conditions, $F(1, 112) = 0.31$, $p = .58$, $\eta 2p < .01$, hemispheres, $F(1, 112) = 0.65$, $p = .42$, $\eta 2p = .01$, or an interaction, $F(1, 112) = 0.81$, $p = .37$, $\eta 2p = .01$. There were no differences in latencies for correct rejections to

high similarity face lures across the stimulation conditions, $F(1, 112) = 0.13$, $p = .72$, $\eta 2p < .01$, hemispheres, $F(1, 112) = 0.21$, $p = .65$, $\eta 2p < .01$, or an interaction, $F(1, 112) = 0.02$, $p = .90$, $\eta 2p < .01$. There were no differences in correct rejections to low similarity face lures across the stimulation conditions, $F(1, 112) = 0.02$, $p = .89$, $\eta 2p < .01$, hemispheres, $F(1, 112) = 0.02$, $p = .90$, $\eta 2p < .00$, or an interaction, $F(1, 112) = 0.57$, $p = .45$, $\eta 2p = .01$. Finally, there were also no differences in correct rejections to new face lures across the stimulation conditions, $F(1, 112) = 0.79$, $p = .38$, $\eta 2p = .01$, hemispheres, $F(1, 112) = 0.01$, $p = .95$, $\eta 2p < .01$, or an interaction, $F(1, 112) = 0.38$, $p = .54$, $\eta 2p < .01$. In sum, there were no differences in latencies as a result of stimulation on the eyewitness identification test.

It is worth noting that caffeine consumption (included as a co-variate in these analyses) had significant effects on response latencies for both hits, $F(1, 112) = 4.75$, $p = .03$, $\eta 2p = .04$, and correct rejections to high similarity face lures, $F(1, 112) = 7.81$, $p = .006$, $\eta 2p = .07$, correct rejections to low similarity face lures, $F(1, 112) = 4.61$, $p = .03$, $\eta 2p = .04$, and new face lures, $F(1, 112) = 5.51$, $p = .03$, $\eta 2p = .04$. As a follow up, we did a post-hoc 2 (prior caffeine consumption: yes or no) x 4 (response latencies to hits, high similarity face lure correct rejections (CRs), low similarity face lure CRs, and new face lure CRs) ANOVA and found a significant main effect of caffeine consumption, $F(1, 117) = 5.39$, $p = .02$, $\eta 2p = .04$, such that participants were *slower* on all trial types if they reporting consuming caffeine in the morning. This suggests participants who drank caffeine before participating in the experiment were more careful in making their judgments on the eyewitness identification task.

Table 5: Mean Proportion of "Yes" Responses on Eyewitness Identification Test as a Function of Brain Stimulation Conditions and Response Latencies (ms) in Experiment 4

	Probability of "yes"			Latencies for Hits/CRs		
	Left	Right	Sham	Left	Right	Sham
Eyewitness Identification Test						
Target	0.60 (.03)	0.60 (.02)	0.63 (.02)	2044	1755	1805
High Similarity Lure	0.38 (.03)	0.41 (.03)	0.44 (.03)	2122	2180	2051
Low Similarity Lure	0.18 (.03)	0.19 (.03)	0.20 (.02)	2012	1860	1903
New	0.14 (.02)	0.17 (.02)	0.16 (.02)	1938	1863	1765

Note. Standard errors of each mean are in parentheses. CRs = correct rejections. Latencies are only for correct responses ("yes" for targets, "no" for lures).

Post-experiment Questionnaire

Although we did not directly ask participants if they thought they were in a stimulation or sham condition, we did ask them if they thought "electricity was being delivered to your brain," if the tDCS procedure had an effect on their performance, and to explain their sensations (see Table 6). The goal of these questions was to determine if there were any differences between the stimulation and sham experiences that might lead to differences in expectations and task performance.

First, we assessed whether participants thought "electricity was being delivered to their brain." There were no differences between left stimulation, right stimulation, and sham participants in their response to this question, $\chi^2 = 0.55$, $p > .76$. The majority of participants believing electricity was being delivered to their brains (86%). When asked whether tDCS increased, decreased, or did not affect their performance, there were also no differences across the three conditions in participant response, $\chi^2 = 4.25$, $p = .37$. The majority of participants responded that stimulation had no effect on their task performance (75%).

When asked to describe their sensations during stimulation, however, participants across the different stimulation conditions responded differently, $\chi^2 = 24.9$, $p < .001$. More specifically, 32% of the left stimulation group and 43% of the right stimulation group reported feeling stimulation throughout the whole 20 minutes of the procedure; however, only 5% of the sham stimulation group reported this. In contrast, 43% of the sham stimulation group reported feeling strong stimulation at the start, but then wondered if stimulation had been turned off for the remaining 20 minutes of the experiment, whereas only 15% of the left stimulation group and 8% of the right stimulation group reported this. This finding is consistent with those of the post-experiment questionnaires in Experiment 2 and 3, which suggested that, when asked, participants are better than chance at indicating whether they received real or sham stimulation.

Table 6: Distribution of reported sensations during tDCS procedure

	Felt stimulation whole time	Strong stimulation first, weaker later	Strong stimulation first, off later	Did not feel stimulation
Left	33%	43%	15%	10%
Right	43%	40%	8%	10%
Sham	5%	48%	43%	5%

Arousal Analyses

In this experiment, we included a pre- and post-stimulation question about self-reported levels of arousal/alertness, asking participants to rate their current arousal level on a 0-9 scale. One common criticism of tDCS is that it does not selectively increase brain function in a target area, but instead boosts activity across the whole brain. This, in turn, might lead to improvements in accuracy or speed on a cognitive task. Although we tried to rule out this possibility in Experiment 1 by including a non-frontal stimulation condition (parietal cortex) to show the specificity of prefrontal stimulation, we nonetheless included this measure to rule out the possibility of overall increases in arousal as a result of stimulating *different areas* of the brain.

The dependent variable used in this analysis for each participant was a post-stimulation arousal rating – pre-stimulation arousal rating difference score. If either of the stimulation conditions leads to overall increases in arousal compared to sham, then this arousal value should be larger in the stimulation conditions than in the sham conditions. The range of these scores were from -5 (5 points lower after stimulation on the scale) to +5 (5 points higher after stimulation on the scale). A one-way ANOVA including left (mean difference = -0.4), right (mean difference = 0), and sham stimulation (mean difference = +0.37) conditions did not find evidence of any difference between the three stimulation conditions, $F(1, 117) = 1.94$, $p = .15$, $\eta 2p = .03$. If anything, stimulation led to numeric *decreases* in overall arousal levels compared to sham.

Combined Analyses

Because we appeared to find a significant effect of stimulation in Experiment 1 and not in Experiments 2-4, we conducted a combined Stimulation x Experiment analysis to determine if the size of the stimulation effect significantly differed across experiments, using the Red Word Test discrimination scores in Experiments 1, 2, and 4 and the green word test discrimination score in Experiment 3 as our dependent measure. A 2 (experiment: 1, 2, 3, 4) x 2 (stimulation condition: left active or sham) ANOVA revealed a trending main effect of stimulation condition, $F(1, 280) = 3.30$, $p = .07$, $\eta 2p = .01$, a main effect of experiment, $F(1, 280) = 9.59$, $p < .01$, $\eta 2p = .10$, and most critically, and a trending interaction, $F(1, 280) = 2.51$, $p = .06$, $\eta 2p = .03$. Although not significant, this trending interaction does suggest that there are, statistically, some potential differences in the size of the stimulation effect across experiments.

Given the time of day effects found in Experiments 2 and 3 that the earliest morning participants had the largest tDCS effects, we split the left stimulation and sham groups into the

earliest morning participants (8 AM and 9 AM) and later morning participants (10 AM and 11 AM). While controlling for arousal, caffeine intake, and sleep quantity and quality, we found a significant interaction, $F(1, 72) = 4.04$, $p = .05$, $\eta 2p = .05$, such that in our early morning participants, stimulation had similar recollection accuracy on the Red Word Test than sham, $t(32) = .84$, $p = .41$, $d = .28$, but on the late morning participants, the sham group had higher recollection accuracy than stimulation, $t(44) = 2.12$, $p = .04$, $d = .63$ (see Table 7). We did not find any significant differences with this breakdown on any of the target or false alarms measures of the eyewitness identification task, all $p > .30$. This is again evidence that stimulation has the largest effects on episodic retrieval in the earliest participants. Doing a combined analysis across Experiments 2-4, we also found this significant interaction, $F(1, 172) = 8.22$, $p < .01$, $\eta 2p = .05$, such that tDCS seemed to "counter" the early morning (stimulation mean: .53 v. sham mean: .39) decreases in the sham condition, but sham participants performed better in the afternoon (stimulation mean: .45 v. sham mean: .51). In general, future tDCS studies should consider time of day as a factor to control.

Given the hint of expectation effects we found in Experiments 2 and 3, we also conducted analyses using only participants who reported feeling the brain stimulation the entirety of the stimulation procedure (i.e., they felt stimulation the "whole time" or they felt "strong stimulation at the start, but weaker stimulation later"). Repeating all of the analyses from above in Experiment 4, but looking only at participants who reported feeling stimulation, we did not find any differences on the criterial recollection test (all $p > .30$) or eyewitness identification test (all $p > .15$) across the stimulation conditions. However, in order to fully explore expectation effects across Experiments 2, 3, and 4, we looked at performance on the word criterial recollection test when pooling the participants who thought they were in the stimulation group

(Experiments 2 and 3) or reported feeling at least some stimulation (Experiment 4). This analysis was intended to remove participants who were suspicious of the sham manipulation. However, in this combined analyses (which included 69 participants in the stimulation group and 43 participants in the sham group), we found no group differences in recollection accuracy (stimulation mean = .48 v. sham mean = .45), $t(110) = .59$, $p = .56$, $d = .11$.

As an additional exploration of expectation effects, we also looked at whether participants thought that stimulation would help, hurt, or have no effect on their performance on the memory test, which was assessed in the post-experiment questionnaires in Experiments 2, 3, and 4. Combing the data from these experiments, we conducted a 3 (expectation effects: help, hurt, or no difference) x 2 (stimulation condition: active or sham) ANOVA using word test accuracy as a dependent variable, but there was no main effect of expectation, $F(1, 210) = .76$, $p = .47$, $\eta 2p = .01$. There was a significant interaction between stimulation condition and expectation effects, $F(2, 210) = 2.98$, $p = .05$, $\eta 2p = .03$, but this was primarily driven by the fact that participants who thought stimulation would have no effect on their performance generally outperformed those who thought it would either help or hurt them in the stimulation condition. Repeating this analysis only with participants who thought stimulation would either help or hurt them, this interaction was no longer significant, $F(1, 44) = .24$, $p = .63$, although it should be noted that this analysis only contained very few participants overall (a total of 48), with extremely uneven distributions of participants in the individual bins (e.g. there were only 6 sham participants who thought stimulation would hurt their performance).

Table 7: Recollection Accuracy on the Red Word Test and Eyewitness Identification Test Broken Down into Early and Late Morning Participants in Experiment 4

	Criterial Recollection Test		Eyewitness Identification Test	
	Left Stimulation	Sham	Left Stimulation	Sham
Early Morning (8 AM, 9 AM)	0.44 (n = 16)	0.38 (n = 18)	0.18 (n = 16)	0.18 (n = 18)
Late Morning (10 AM, 11 AM)	0.38 (n = 24)	0.51 (n = 22)	0.25 (n = 24)	0.20 (n = 22)

Note. The recollection accuracy score for the Eyewitness Identification Test was computed by subtracting false alarms to high similarity faces from hits to target faces. It is presented this way for simplicity, since none of the four measures (hits to targets, false alarms to high similarity faces, false alarms to low similarity faces, or false alarms to new faces) differed as a result of time of stimulation.

Discussion

Contrary to our expectations, we were again unable to replicate the finding that electrically stimulating prefrontal cortex leads to improvements in retrieval accuracy. This is particularly surprising given the task had identical instructions to that of Experiment 1 and that this experiment had significantly increased power from the previous experiments. We also did not find any evidence that either left or right prefrontal stimulation led to improvements in accuracy or speed of eyewitness identification judgments. This lack of replicability is consistent with the concerns of several critics of the use of tDCS to manipulate frontal brain activity (Tremblay et al., 2014; Jacobson, Koslowsky, & Lavidor, 2011; Brunoni & Vanderhasselt, 2014).

Before drawing conclusions, it is important to note the differences between Experiments 1 and 4. In Experiment 1, tDCS was applied in a separate room and using a different machine than that used in Experiments 2, 3, and 4. Furthermore, in Experiment 4, although the task instructions and most of the stimuli were identical as those in Experiment 1, we only included about half the total number of stimuli, and as such, total memory load was decreased for the

criterial recollection task. The reason that we did this is because we also included an eyewitness identification task, which contained different instructions and an entirely new set of stimuli to make up for this decrease in memory load. Still, given the similarity of this task to not only that of Experiment 1 but also the task used in Experiments 2 and 3, we would expect that the effects of stimulating and increasing prefrontal activity would not be so task specific.

It is difficult to draw conclusions about why we were unable to replicate the improvements in retrieval accuracy from Experiment 1. We know from the abundance of previous neuroimaging studies using source monitoring tasks (see "Biological Correlates of Retrieval" in the Introduction) that memory retrieval for specific details is correlated with frontal activity. However, it may be the case that this activity is confounded with factors such as task difficulty. For example, perhaps the correlation between dlPFC and retrieval monitoring in the neuroimaging literature is a result of task difficulty, such that dlPFC becomes more active for more challenging memory trials that require more monitoring. Maybe dlPFC does not contribute directly to the monitoring process, but instead acts as an indicator that a particular trial is more challenging and consequently increases selective attention to that trial. In any case, while we did find evidence in Experiment 1 that dlPFC stimulation led to increased recollection accuracy, we were unable to replicate this finding here with a larger sample and are thus left with little evidence to suggest dlPFC is causally related to episodic retrieval processes.

As for the eyewitness identification data, we were unable to find evidence that increasing the underlying activity in left or right dlPFC led to changes in face identification accuracy, consistent with the anodal part of the Smirni and colleagues (2015) experiment that used a three-alternative forced choice (i.e., simultaneous presentation) paradigm. Although we might have expected to see frontal activity when participants were making demanding judgments as to

whether or not they had studied the face before (as was suggested by Rapcsak, 2001 and Elgar & Campbell, 2001), this study was largely exploratory, as there have not been many studies looking specifically at the retrieval network when identifying face stimuli. Perhaps dlPFC and other prefrontal regions have minimal involvement in discriminating between similar face stimuli, and these tasks are instead dependent on the parietal and visual areas seen in typical face recognition studies.

Although we did not find memory effects as a result of prefrontal stimulation, our post-experiment questionnaire revealed that participants were able to discern differences in physical sensations between the stimulation and sham procedures when prompted. More specifically, nearly half of the sham participants indicated that they "wondered if stimulation had been turned off," whereas only 11% of participants actually being stimulated (combined left and right) felt this way. We re-analyzed all of the results using only participants who reported feeling at least some stimulation for the entire procedure, but did not find any significant memory effects, unlike in Experiments 2 and 3. Still, it is important to consider that sensations from tDCS may lead some participants (such as those that have been in another tDCS study before) to discern whether or not they are in a sham condition.

GENERAL DISCUSSION

The present research investigated the effects of stimulating left and right dlPFC using anodal tDCS on episodic memory retrieval. Previous studies using tDCS have generally shown that anodal stimulation of prefrontal cortex increases working memory performance (see Kuo & Nitsche, 2012 for review), but results of the research on episodic memory have been inconsistent. In our experiments, participants received either active or sham tDCS over left or right dlPFC after studying a series of items for a memory test. Stimulation was applied

immediately before retrieval. This is distinct from the majority of brain stimulation studies, which typically apply stimulation before encoding (see Manenti, Cotelli, Robertson, & Miniussi, 2012 for review). Applying stimulation before encoding has the disadvantage of affecting both encoding and retrieval processes, at least in studies in which the memory test occurs immediately after the study phase. With our study design, we aimed to demonstrate a causal role of left or right prefrontal cortex regions (dlPFC in particular) specific to retrieval.

In Experiment 1, we applied either active or sham stimulation to left or right dlPFC immediately before participants completed a criterial recollection test, which tested their ability to remember details about words and pictures they had studied earlier. The main finding was that participants who received left stimulation, and to a smaller extent right stimulation, were better able to remember details about word stimuli than participants who received sham stimulation. In Experiments 2 and 3, we included versions of the criterial recollection task that were either easier or more demanding than Experiment 1, respectively. In both of these experiments, we were unable to replicate the significant benefits of left stimulation over dlPFC on recollection accuracy. Experiment 4 was designed to provide a replication attempt of Experiment 1 with greater power. A secondary goal of Experiment 4 was to explore if increasing the activity in dlPFC via stimulation could lead to improvements on an eyewitness identification memory task using face stimuli. We were again unable to replicate the stimulation effects from Experiment 1, and found no benefits of stimulation on the eyewitness memory task.

Although we found evidence that dlPFC can contribute to episodic memory retrieval in Experiment 1, we were unable to replicate this finding in Experiments 2, 3, and 4. Post-hoc analyses suggested that the time of day of stimulation mattered, such that our early morning stimulation participants generally outperformed our sham stimulation participants. These null

results cloud our interpretation of Experiment 1 that dlPFC plays a causal role in episodic retrieval. Because of the lack of replicability, a number of questions are left unresolved: does dlPFC reliably contribute to episodic memory retrieval? If so, how do left and right dlPFC each separately contribute to these retrieval processes? Are there boundary conditions in which dlPFC does contribute to retrieval, or are the findings in the neuroimaging literature simply a correlational artifact?

Theoretical Implications

When trying to recall memorial information about a cue word during retrieval, there are two major processes that take place: memory search and postretrieval monitoring. Memory search is the process through which one attempts to actively bring an existing memory trace to mind. Postretrieval monitoring describes the decision processes through which one analyzes the memorial evidence attained via the search to determine whether the memorial information attained is valid. In these experiments, tDCS was applied immediately before a retrieval test, meaning that either of these processes could have been affected by stimulation.

Based on the previous neuroimaging literature, we had hypothesized that anodal stimulation over dlPFC would be most likely to affect postretrieval monitoring processes. We expected that left dlPFC stimulation would lead to the most pronounced effects on disqualifying monitoring and that right dlPFC stimulation would lead to the most pronounced effects on diagnostic monitoring. We found no effects of stimulation on the disqualifying monitoring test in Experiment 1. We also found little evidence across the 4 experiments to suggest that diagnostic monitoring was improved as a result of left or right anodal stimulation. In Experiment 1, while we did find that dlPFC contributed to recollection accuracy, this effect was primarily driven by memory for target information, as opposed to improvements in false alarm levels to critical lures.

This suggests the role of left dlPFC in retrieval is, if anything, in memory search. However, we were unable to replicate this finding across the three remaining studies.

There are a multitude of potential explanations that could explain the null results of Experiments 2-4, but it is impossible to determine which is true without further research. First, given the uncertainty of the efficacy of tDCS in its application to dlPFC, we do not know if our null results were because the tDCS procedure had limited or no effects with our particular stimulation montage in these experiments. Second, we do not know if any of the minor changes in task or stimulation protocols in these experiments somehow interacted with the size of the stimulation effect. In order to make our interpretations in the context of the existing literature, we will assume for the rest of this discussion that application of tDCS worked as intended; however, it is important to keep in mind that this assumption may be unwarranted.

Given the inconsistency across these three experiments, and the fact that Experiment 4 showed a complete lack of effects of anodal dlPFC stimulation despite a substantially increased number of subjects and better control over several confounding factors, we are left to conclude that increasing dlPFC activity does not affect episodic retrieval in a reliable way. Our null results contradict the claims of the neuroimaging literature, which suggest a critical role of prefrontal regions for executive control of retrieval monitoring (e.g., Gallo et al., 2013, Mitchell et al., 2008). It is important to consider that the effects in these neuroimaging studies are correlational. For example, Gallo and colleagues (2013) compared brain activity between memory for words and pictures using a criterial recollection task identical to that used in Experiment 1. When the authors contrasted correct rejections for noncriterial lures on the word and Picture Tests, they found evidence of increased prefrontal activity on the Word Test that they attributed to more effortful retrieval monitoring processes. However, the effects are not necessarily specific to

retrieval monitoring; it may simply be the case that dlPFC is more active for the most challenging items (i.e., the more challenging red word trials), perhaps as a result of increased selective attention to the most challenging items. The idea that the Word Test was more challenging than the Picture Test is consistent with both our and Gallo and colleagues' latency data for the task, which showed that subjects were generally slower when judging noncriterial lures on the Word Test than on the Picture Test. In the context of our experiments, perhaps increasing selective attention to the most challenging stimuli via tDCS over dlPFC was not sufficient to produce increases in memory performance.

We also found little evidence of laterality differences. In Experiment 1, while recollection accuracy effects were strongest in left dlPFC, there also appeared to still be a moderate effect of stimulating right prefrontal cortex on recollection accuracy, both of which were driven by increased memory for targets. We did not apply stimulation to right dlPFC in Experiments 2 and 3. In Experiment 4, we were unable to find evidence of a stimulation effect in either hemisphere. Consequently, given the lack of replicability across Experiments 1 and 4, if there were any effects of dlPFC laterality, they were not reliable.

The purpose of using brain stimulation methods like tDCS is to supplement the correlational limitations inherent in neuroimaging designs. Although it is often tempting to draw conclusions about cognitive processes using neuroimaging contrasts, there are always a large number of confounding cognitive processes that we cannot always easily control, such as task difficulty, attention, or item salience to participants. By combining the results of neuroimaging and brain stimulation techniques together, however, we can get around some of these limitations. In this case, we have discovered that although dlPFC activity seemed to be correlated with more effortful retrieval monitoring in the neuroimaging literature, tDCS over this region was not

sufficient to reliably improve behavioral measures of retrieval monitoring or recollection accuracy in general.

Limitations

Although we carefully controlled for a large number of factors across the four experiments, there were a number of limitations that may cloud the interpretation of our results. First, although we applied stimulation to dlPFC immediately before retrieval, it is possible we also affected memory consolidation processes during the stimulation time period. A number of studies (Sandrini et al., 2014; Sandrini, Censor, Mishoe, & Cohen, 2013; Javadi & Cheng, 2013; Marshall, Molle, Hallschmid, & Born, 2004) suggest that applying tDCS after study can affect how memories are consolidated, which then has long-term effects on the quantity of information people can remember. Although we were trying to specifically target monitoring processes, the slow effects of tDCS required us to apply stimulation before retrieval began. The consolidation account could also explain the improvements in recollection accuracy found exclusively in Experiment 1. Because these effects were driven by memory for target information, perhaps stimulating dlPFC enhanced consolidation of the studied items during the 30-minute time interval between study and test. In fact, specific to Experiment 1, there was, on average, about 10 minutes more between study and test than in Experiments 2-4. If our effects were driven by enhancing consolidation rather than memory search, this may explain why we were unable to replicate the results of Experiment 1 in the other experiments.

As mentioned above, there was more time between stimulation and test in Experiment 1 than in Experiments 2-4. Experiment 1 was also conducted in a single blind design, whereas Experiments 2-4 were semi-double blinded. Furthermore, Experiment 1 was conducted before we attained our own tDCS machine in the lab. Consequently, the stimulation procedure in

Experiment 1 occurred in a different lab space. This also meant that we used a different tDCS machine in Experiment 1 than in Experiments 2-4. We also used a different consent form in Experiment 1 than in Experiments 2 and 3. Any of these differences may explain why we were unable to replicate the results of Experiment 1.

Methodological Considerations for Future tDCS Studies

A general concern in using tDCS on cortical regions to modulate cognitive performance is the general lack of replicability across similar studies, as was the case in this series of experiments. Many of the assumptions and parameters used in cognitive tDCS studies are derived from studies applying stimulation to motor cortex to modulate motor evoked potentials. (e.g., Nitsche & Paulus, 2000). Some of these assumptions, such as the importance of electrode polarity, may be unwarranted for use in other cortical regions. Furthermore, some reviews and meta-analyses of papers using tDCS to modulate various cognitive domains such as working memory suggest the technique has inconsistent effects at best (Tremblay et al., 2014; Jacobson, Koslowsky, & Lavidor, 2011; Brunoni & Vanderhasselt, 2014). Berryhill and colleagues (2014) conducted three studies exploring the use of anodal tDCS over dlPFC to facilitate working memory function, and were also unable to replicate their results. The authors of this study suggest that tDCS studies are often limited by insufficiently challenging tasks, patient populations with high levels of heterogeneity, and too few participants. Tremblay and colleagues' (2014) review corroborates the latter point, as the mean sample size of the literature reviewed was only 21 participants per condition. A power analysis suggests, even assuming a large effect size (defined as Cohen's d = .80, which is much larger than typical tDCS effects), a properly powered study would require 26 participants in each group. As such, it is critical that

future studies using tDCS to explore cognitive domains are done with large enough sample sizes for the study to be sufficiently powered (as we were careful to do in Experiment 4).

In addition to these factors, there are several other concerns about tDCS research that were raised by this research. One major cause for concern in this study was the importance of using a double blinding procedure. Experiment 1 was done in a single blind fashion in which the participant was not informed whether they were receiving active or sham stimulation, but the experimenter knew the entire time. This design is consistent with the majority of tDCS studies in the literature. The reason this design is typical is because on most commercial stimulators, the experimenter has to physically flip a switch that says "sham," so they must be aware of the stimulation condition before actually starting the stimulator. To get around this in Experiments 2-4, we performed a semi-double blind procedure in which the experimenters set up the stimulation apparatus and provided all instructions before knowing whether they needed to set the stimulator to sham. Then, at the moment immediately before stimulation began, the experimenter checked whether or not to hit the sham button. Although this method is not a perfect double blinding procedure, it prevented the experimenter from providing the stimulation instructions in a biased manner. The fact that we could not replicate the results of Experiment 1 after we introduced this blinding procedure suggests it may be a potential source of inconsistency in study design that leads to Type I error in interpreting the results of tDCS studies (i.e., the "Clever Hans" effect, Pfungst, 1911).

Another concern raised by these experiments was evidence that participants could, on average, tell when they were in the sham stimulation condition. In Experiments 2 and 3, we explicitly asked participants, after completing the study, if they thought they received active stimulation. Assuming that the sham procedure was convincing, there should have been no

difference between the active and sham stimulation groups' response to this question. However, in both experiments, participants in the sham group were more likely than stimulation participants to indicate they were not receiving active stimulation. To explore this finding further, we asked participants in Experiment 4 to describe their sensations with one of four responses: 1) felt stimulation the entire time, 2) felt strong stimulation at first and then weaker stimulation, 3) felt strong stimulation at first and then no stimulation and thought the current had been turned off, or 4) did not feel stimulation at all. Consistent with Experiments 2 and 3, participants in the sham condition were more likely to indicate that the current had turned off after the initial, brief stimulation period. In general, this is a worrying trend that may lead some participants (particularly those who have been in tDCS studies before) to know when they are not receiving active stimulation. This could, for example, lead to unintended expectation effects that could affect participants' motivation to take the study tasks seriously. In general, to control for this in future studies, tDCS researchers should consider 1) asking participants whether they have participated in another tDCS study before and 2) at the end of the study, asking participants if they suspected that they were in a sham condition.

An interesting trend within this series of experiments was the time of day effects. In Experiments 2-4, there was evidence of stimulation differences in recollection accuracy based on whether participants were run early in the morning or later in the day. In Experiments 2 and 3, this effect was driven by the fact that stimulation participants outperformed sham participants in the morning, but not in the afternoon. In Experiment 4, it was driven by the fact that sham participants in the late morning drastically outperformed their stimulation counterparts, whereas there were no group differences in the early morning. It could be the case that bringing the sham participants down to a "lower" level of cognitive function created a sufficiently challenging

scenario such that tDCS could have effects (akin to the ideas of Berryhill and colleagues, 2014) in the early morning, at least in Experiments 2 and 3. It may also be the case that active stimulation made morning participants more "motivated" to try on the memory test, whereas sham stimulation did not have such effects. This is a question that warrants future research, but in general, researchers conducting tDCS studies should consider the importance of time of day when planning participant recruitment.

Future Studies

Although we initially showed evidence that might suggest dlPFC has a causal role in episodic memory retrieval, we were unable to replicate these results across the three follow-up experiments. However, this does not mean that dlPFC is not involved in episodic recollection overall – just that tDCS was not sufficient to improve recollection accuracy in a criterial recollection task. There are still very few studies that apply stimulation before or during retrieval on any episodic memory task. As such, future research should consider stimulating dlPFC and perhaps other regions (such as temporoparietal lobe) before retrieval in other episodic memory paradigms. It would be interesting, for example, to explore if brain stimulation before retrieval can differentially affect the use of recollection-based and familiarity-based retrieval processes (see Yonelinas, 2002 for review).

An exciting branch of research that has recently begun is the use of tDCS in older adult populations, particularly those with dementia, to help combat memory loss. Some have suggested that brain stimulation techniques may have potential for improving symptoms of dementia (see Liao and colleagues, 2015 or Elder & Taylor, 2014 for review). Less research, however, has been done with healthy older adult populations. Given the hypothesis of a posterior-anterior shift in neural activation for older adults (Davis et al., 2008), which suggests

older adults become more dependent on frontal processes during memory retrieval as their brains lose access to the more detailed information associated with recollection, it might be interesting to apply stimulation to sensory regions before or during retrieval to "compensate" for these deficits. One could also apply stimulation to frontal regions to see if it further facilitates memory accuracy. Such research would serve as a useful complement to the large body of neuroimaging literature about the aging brain.

Although we did not find evidence of improved eyewitness identification accuracy as a result of frontal stimulation, there are a number of unanswered questions that may be resolved through the use of brain stimulation techniques. Visual regions typically involved in the face recognition process, such as the fusiform gyrus, are good targets for future stimulation research. Other studies may also want to consider using more realistic stimuli, such as videos of crime scenes, during encoding. Although we did have participants study multiple faces, these faces were only sketches. Furthermore, witnessing a real crime scene (perhaps in a video) would have more of an emotional impact on the participant, which could then affect the neural processes involved in face recognition during an eyewitness identification task. Stimulation studies using video crime scene stimuli should thus consider targeting emotional regions, such as the cortical network connected to the amygdala.

There is also a need for more episodic memory research using other brain stimulation techniques in general. Given that tDCS applies a weak amount of brain stimulation in comparison to TMS and there appeared to be a small effect of dlPFC stimulation on memory recollection accuracy, future research should consider applying TMS to dlPFC on source memory tasks. This may more strongly increase activity in the region by directly inducing neural firing and could lead to more replicable findings across experiments. It may also be the case that

increasing activity in dlPFC would not boost memory accuracy further, but dlPFC may still be involved in some capacity. As such, future researchers should also consider applying TMS or cathodal tDCS to temporarily *lesion* dlPFC in order to see if memory recollection accuracy decreases.

Future research should also consider high definition tDCS (Kuo et al., 2013) for more precise stimulation. One major disadvantage of the sponge tDCS technique used in these experiments is that the current is widely distributed between the anode and the cathode. As such, frontal regions other than dlPFC received current as well, and the activation of such regions (should they be inhibitory to dlPFC) may have interfered with our manipulation. HD-tDCS gets around this problem by using a smaller anode that is surrounded by a ring of cathodes located directly over the targeted region. A disadvantage of this technique is that the current penetrates less deeply into cortex than in sponge tDCS. There also simply has not been much research done with the technique at this point in time, especially in episodic memory. The only published episodic memory study using HD-tDCS is by Nikolin and colleagues (2015), who found evidence that HD-tDCS applied over dlPFC, the planum temporale, and the left medial temporal lobe (all regions functionally connected to hippocampus) can increase the rate of verbal learning and the speed of working memory performance, although they did not find differences in accuracy. More research is certainly needed using this exciting new technique.

Future research must also combine tDCS (and HD-tDCS) with neuroimaging methods more directly. There are now a number of commercial tDCS machines available that are compatible with MR scanners. Combining tDCS with fMRI/ERP data has a number of advantages. First, the researcher has better insight into where the current is going during stimulation through the use of functional neuroimaging techniques. It also allows for the use of

advanced techniques like functional connectivity analyses. It could be the case that increased excitability in a targeted brain region with tDCS also has downstream effects on functionally connected regions; functional connectivity analyses can help determine whether this is the case. Only by converging methods can researchers truly understand how stimulation methods like tDCS are affecting brain function.

Conclusions

This was the most extensive series of experiments to use tDCS to explore the role of prefrontal cortex in episodic memory retrieval. Although we were unable to replicate the behavioral results of our first experiment in the a priori whole sample analyses of the subsequent experiments, we did find some evidence for early morning effects. However, more work needed to determine the reliability of these morning effects. The lack of reliable tDCS effects is still informative in the context of previous studies. This research provides new questions about the role of prefrontal cortex in episodic memory and methodological insights for future studies that utilize tDCS and other brain stimulation technologies. Although many recent media reports on tDCS make it out to be some kind of miracle intervention in which you "upload data to your brain at a quicker rate" like in the movie *The Matrix* (Morales, 2016), our results suggest we are a long way from creating another Neo. If tDCS over dlPFC does affect the amount of detail information you can remember, the effects are subtle and not easily replicable across different memory tasks.

REFERENCES

Bakker, A., Kirwin, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science, 319,* 1640-1642.

Barbosa, F. F. & Albuquerque, F. S. (2008). Effect of time-of-day of training on explicit memory. *Brazilian Journal of Medical and Biological Research, 41,* 477-481.

Barker, A. T., Jalinous, R., & Freeston, I. L. (1985). Non-invasive magnetic stimulation of human motor cortex. *Lancet, 2,* 1106-1107.

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*, 800.

Blanchet, S., Gagnon, G., & Schneider, C. (2010). The contribution of the dorsolateral prefrontal cortex in full and divided encoding: A paired-pulse transcranial magnetic stimulation study. *Behavioral Neurology, 23,* 107-115.

Blumenfeld, R. S., Lee, T., & D'Esposito, M. (2014). The effects of lateral prefrontal transcranial magnetic stimulation on item memory encoding. *Neuropsychologia, 53,* 197-202.

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

Bona, S. & Silvanto, J. (2014) Accuracy and confidence of visual short-term memory do not go hand-in-hand: Behavioral and neural dissociations. *PLoS One, 9,* e90808.

Brunoni, A. R. & Vanderhasselt, M. (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.

Buckner, R. L. & Koutstaal, W. (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proceedings of the National Academy of Science, USA, 95,* 891-898.

Buckner, R. L., Wheeler, M. E., & Sheridan, M. A. (2001). Encoding processes during retrieval tasks. *Journal of Cognitive Neuroscience, 13,* 406-415.

Cabeza, R., Locantore, J. K., & Anderson, N. D. (2003). Lateralization of prefrontal activity during episodic memory retrieval: Evidence for the production-monitoring hypothesis. *Journal of Cognitive Neuroscience, 15,* 249-259.

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.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex, 12,* 1048–1056.

Chi, R. P., Fregni, F., & Snyder, A. W. (2010). Visual memory improved by non-invasive brain stimulation. *Brain Research, 1353,* 168-175.

Chua, E. F. & Ahmed, R. (2016). Electrical stimulation of the dorsolateral prefrontal cortex improves memory monitoring. *Neuropsychologia, 85,* 74-79.

Clark, D. & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia, 41,* 304-317.

Cruse, D. & Wilding, E. (2009). Prefrontal cortex contributions to episodic retrieval monitoring and evaluation. *Neuropsychologia, 47,* 2779-2789.

Davis, N. J., Gold, E., Pascual-Leone, A., & Bracewell, R. M. (2013). Challenges of proper placebo control for non-invasive brain stimulation in clinical and experimental applications. *European Journal of Neuroscience, 38,* 2973-2977.

Davis, S. W., Dennis, N. A., Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2008). Qué PASA? The Posterior-Anterior Shift in Aging. *Cerebral Cortex, 18,* 1201-1209.

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron, 35,* 989-996.

Duecker, F., & Sack, A. T. (2013). Pre-Stimulus sham TMS facilitates target detection. *PLoS One, 8,* e57765.

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

Elder, G. J. & Taylor, J. P. (2014). Transcranial magnetic stimulation and transcranial direct current stimulation: Treatments for cognitive and neuropsychiatric symptoms in the neurodegenerative dementias? *Alzheimer's Research Therapy, 6,* 74.

Elgar, K. & Campbell, R. (2001). Annotation: The cognitive neuroscience of face recognition: Implications for developmental disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 42,* 705-717.

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,* 1-7.

Fell, J., Staresina, B. P., Do Lam, A. T. A., Widman, G., Helmstaedter, C., Elger, C. E., & Axmacher, N. (2013). Memory modulation by weak synchronous deep brain stimulation: A pilot study. *Brain Stimulation, 6,* 270-273.

Finley, J. R., Roediger, H. L. R., Hughes, A. D., Wahlheim, C. N., & Jacoby, L. L. (2015). Simultaneous versus sequential presentation in testing recognition memory for faces. *The American Journal of Psychology, 128,* 173-195.

Fregni F, Boggio P. S., Nitsche M., Bermpohl F., Antal A., Feredoes E., Marcolin M. A., Rigonatti S. P., Silva M. T., Paulus W., & Pascual-Leone A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research, 166,* 23-30.

Gagnon, G., Blanchet, S., Grondin, S., & Schneider, C. (2010). Paired-pulse transcranial magnetic stimulation over the dorsolateral prefrontal cortex interferes with episodic encoding and retrieval for both verbal and non-verbal materials. *Brain Research, 1344,* 148-158.

Gagnon, G., Schneider, C., Grondin, S., & Blanchet, S. (2011). Enhancement of episodic memory in young and healthy adults: A paired-pulse TMS study on encoding and retrieval performance. *Neuroscience Letters, 2,* 138-142.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49,* 87-115.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 120-128.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38,* 833-848.

Gallo, D. A. (2013). Retrieval expectations affect false recollection: Insights from a criterial recollection task. *Current Directions in Psychological Science, 22,* 316-323.

Gallo, D. A., Foster, K. T., & Johnson, E. L. (2009). Elevated false recollection of emotional pictures in young and older adults. *Psychology and Aging, 24,* 981-988.

Gallo, D.A., McDonough, I.M. & Scimeca, J.M (2010). Dissociating source memory decisions in the prefrontal cortex: fMRI of diagnostic and disqualifying monitoring. *Journal of Cognitive Neuroscience, 22,* 955-969.

Gallo, D. A., Kensinger, E. A., & Schacter, D. L. (2006). Prefrontal activity and diagnostic monitoring of memory retrieval: fMRI of the criterial recollection task. *Journal of Cognitive Neuroscience, 18,* 135-148.

Gallo, D. A., Weiss, J.A., & Schacter, D.L. (2004). Reducing false recognition with criterial recollection tasks: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473-493.

Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., Kaelin-Lang, A., Mima, T., Rossi, S., Thickbroom, G. W., Rossini, P. M., Ziemann, U., Valls-Solé, J., & Siebner, H. R. (2012). A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clinical Neurophysiology, 123,* 858-882.

Halbig, T. D., Gruber, D., Kopp, U. A., Scherer, P., Schneider, G. H., Trottenberg, T., Arnold, G., & Kupsch, A. (2004). Subthalamic stimulation differentially modulates declarative and nondeclarative memory. *Neuroreport, 15,* 539-543.

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature, 406,* 147-150.

Hawco, C., Berlim, M. T., & Lepage, M. (2014). The dorsolateral prefrontal cortex plays a role in self-initiated elaborative cognitive processing during episodic memory encoding: rTMS evidence. *PLoS One, 8,* e73789.

Hayama, H. R., & Rugg, M. D. (2009). Right dorsolateral prefrontal cortex is engaged during post-retrieval processing of both episodic and semantic information. *Neuropsychologia, 47,* 2409-2416.

Henson, R. N. A., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: A functional MRI test of the monitoring hypothesis. *Brain, 122,* 1367–1381.

Herron J. E. & Rugg M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin and Review, 10,* 703-710.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20,* 167–173

Iidaka, T., Harada, T., Kawaguchi, J., & Sadato, N. (2012). Neuroanatomical substrates involved in true and false memories for face. *Neuroimage, 62,* 167-176.

Jacobson, L., Goren, N., Lavidor, M., & Levy, D. A. (2012). Oppositional transcranial direct current stimulation (tDCS) of parietal substrates of attention during encoding modulates episodic memory. *Brain Research, 1439,* 66-72.

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

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383-400). New York: Guilford Press.

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology, 10,* 370-375.

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

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

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

Johnson, M. K. (2006). Memory and Reality. *American Psychologist, 61,* 760-771.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28.

Jones, K. T., Gozenman, F., & Berryhill, M. E. (2014). Enhanced long-term memory encoding after parietal neurostimulation. *Experimental Brain Research, 232,* 4043-4054.

Jones, K. T., Gozenman, 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.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51,* 481-537.

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,* 644-648.

Kuo, M. F. & Nitsche, M. A. (2012). Effects of transcranial electrical stimulation on cognition. *Clinical EEG and Neuroscience, 43,* 192-199.

Kuo, M. F. & Nitsche, M. A. (2015). Exploring prefrontal cortex functions in healthy humans by transcranial electrical stimulation. *Neuroscience Bulletin, 31,* 198-206.

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. & Stark, C. E. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory, 18,* 15-18.

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology, 44,* 894-904.

Lepage, M. (2004). Differential contribution of left and right prefrontal cortex to associative cued-recall memory: A parametric PET study. *Neuroscience Research, 48,* 297–304.

Liao, X., Li, G. M., Wang, A. G., Liu, T., Feng, S. G., Guo, Z. W., Tang, Q., Jin, Y., Xing, G. Q., McClure, M. A., Chen, H. P., He, B., Liu, H., & Mu, Q. W. (2015). Repetitive transcranial magnetic stimulation as an alternative therapy for cognitive impairment in Alzheimer's Disease: A meta-analysis. *Journal of Alzheimer's Disease, 48,* 463-472.

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,* 372-380.

Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. *Current Biology, 25,* 780-785.

Luber, B. & Lisanby, S. H. (2014). Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *Neuroimage, 85,* 961-970.

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.

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,* 3345-3351.

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,* UNSP 49.

Manenti, R., Cotelli, M., Calabria, M., Maioli, C., & Miniussi, C. (2010). The role of the dorsolateral prefrontal cortex in retrieval from long-term memory depends on strategies: A repetitive transcranial magnetic stimulation study. *Neuroscience, 166,* 501-507.

Manenti, R., Tettamanti, M., Cotelli, M., Miniussi, C., & Cappa, S. F. (2010). The neural bases of word encoding and retrieval: A fMRI-guided transcranial magnetic stimulation study. *Brain Topography, 22,* 318-332.

Manenti, R., Cotelli, M., Robertson, I. H., & Miniussi, C. (2012). Transcranial brain stimulation studies of episodic memory in young adults, elderly adults and individuals with memory dysfunction: A review. *Brain Stimulation, 5,* 103-109.

Marshall, L., Molle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *The Journal of Neuroscience, 24,* 9985-9992.

McDonough, I. M., Wong, J. T., & Gallo, D. A. (2012). Age-related differences in prefrontal cortex activity during retrieval monitoring: Testing the compensation and dysfunction accounts. *Cerebral Cortex, 23,* 1049-1060.

Merton, P. A. & Morton, H. B. (1980). Stimulation of the cerebral cortex in the intact human subject. *Nature, 285,* 227.

Mitchell, K. J. & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135,* 638-677.

Mitchell, K. J., Raye, C. L., McGuire, J. T., Frankel, H., Greene, E. J., & Johnson, M. K. (2008). Neuroimaging evidence for agenda-dependent monitoring of different features during short-term source memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 780–790.

Morales, J. R. (March 8, 2016). Learn new skills: Can brain uploads work for real? *HealthAim.* Retrieved from http://www.healthaim.com/learn-new-skills-can-brain-uploads-work-real/44432

Morgan, H. M., David. N. J., & Bracewell, R. M. (2014). Does transcranial direct current stimulation affect mood and emotional memory retrieval in healthy individuals? *PLoS One, 9,* e92162.

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,* 2384-2389.

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.

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

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

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

Nitsche, M. A., Seeber, A., Frommann, K., Kelin, C. C., Rochford, c., Nitsche, M. S., Fricke, K., Liebetanz, D., Lang, N., Antal, A., Paulus, W., & Tergau, F. (2005). Modulating parameters of excitability during and after transcranial direct current stimulation of the human motor cortex. *Journal of Physiology, 568,* 291-303.

Nolde, S. F., Johnson, M. K., & Raye, C. L. (1998). The role of prefrontal cortex during tests of episodic memory. *Trends in Cognitive Sciences, 2,* 399-406.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97-113.

Penolazzi, B., Di Domenico, A., Marzoli, D., Mammarella, N., Fairfield, B., Franciotti, R., Brancucci, A., & Tommasi, L. (2010). Effects of transcranial direct current stimulation on episodic memory related to emotional visual stimuli. *PLoS One, 5,* e10623.

Penolazzi, B., Stamaccia, D. F., Braga, M., Mondini, S., & Galfano, G. (2014). Human memory retrieval and inhibitory control in the brain: Beyond correlational evidence. *Journal of Neuroscience, 34,* 6606-6610.

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

Pfungst, O. (1911). Clever Hans (the horse of Mr. von Osten): A contribution to experimental animal and human psychology (Trans. C. L. Rahn). New York: Henry Holt. (Originally published in German, 1907).

Pisoni, A., Turi, Z., Raithel, A., Ambrus, G. C., Alekseichuk, I., Schact, A., Paulus, W., & 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,* e012385.

Poreisz, C., Boros, K., Antal, A., & Paulus, W. (2007). Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. *Brain Research Bulletin, 72,* 208-214.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2000). Left anterior prefrontal activation increases with demands to recall specific perceptual information. *The Journal of Neuroscience, 20,* RC108.

Ranganath, C. & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience, 13,* 713-726.

Rapcsak, S. Z., Nielsen, L., Littrell, L. D., Glisky, E. L, Kaszniak, A. W., & Laguna, J. F. (2001). Face memory impairments in patients with frontal lobe damage. *Neurology, 57,* 1168-1175.

Reyna, V. F. & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9,* 95-123.

Richmond, L. L., Wolk, D., Chein, J., & Olson, I. R. (2014). Transcranial direct current stimulation enhances verbal working memory training performance over time and near transfer outcomes. *Journal of Cognitive Neuroscience, 26,* 2443-2454.

Ridding, M. C. & Ziemann, U. (2010). Determinants of the induction of cortical plasticity by non-invasive brain stimulation in healthy subjects. *Journal of Physiology, 588,* 2291-2304.

Roediger, H. L. III & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 814-816.

Roediger, H. L. III & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9,* 123-127.

Roediger, H. L. III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review,* 8, 385-407.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & The Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology, 120*, 2008–2039.

Rossi, S., Miniussi, C., Pasqualetti, P., Babiloni, C., Rossini, P. M., & Cappa, S. F. (2004). Age-related functional changes of prefrontal cortex in long-term memory: a repetitive transcranial magnetic stimulation study. *Journal of Neuroscience, 24,* 7939-7944.

Rossi, S., Pasqualetti, P., Zito, G., Vecchio, F., Cappa, S. F., Miniussi, C., Babiloni, C., & Rossini, P. M. (2006). Prefrontal and parietal cortex in human episodic memory: An interference study by repetitive transcranial magnetic stimulation. *European Journal of Neuroscience, 23,* 793-800.

Rugg, M. D., Henson, R. N., & Robb, W. G. (2003). Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia, 41,* 40-52.

Rugg M. D., Johnson J. D., Park H. K., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Progress in Brain Research, 169,* 339-352.

Rugg, M.D. & Vilberg, K.L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology, 23,* 255-260.

Rylas, A. J., Rogers, L. M., Gross, E. Z., Polnaszek, K. L., & Voss, J. L. (2016). Associative recognition memory awareness improved by theta-burst stimulation of frontopolar cortex. *Cerebral Cortex, 26,* 1200-1210.

Sandrini, M., Brambilla, M., Manenti, R., Rosini, S., Cohen, L. G., & Cotelli, M. (2014). Noninvasive stimulation of prefrontal cortex strengthens existing episodic memories and reduces forgetting in the elderly. *Frontiers in Aging Neuroscience, 6,* 289.

Sandrini, M., Censor, N., Mishoe, J., & Cohen, L. G. (2013). Causal role of prefrontal cortex in strengthening of episodic memories through reconsolidation. *Current Biology, 23,* 2181-2184.

Sandrini, M., Manenti, R., Brambilla, M., Cobelli, C., Cohen, L. G., & Cotelli, M. (2016). Older adults get episodic memory boosting from noninvasive stimulation of prefrontal cortex during learning. *Neurobiology of Aging, 39,* 210-216.

Sestieri, C., Capotosto, P., Tosoni, A., Luca Romani, G., & Corbetta, M. (2013). Interference with episodic memory retrieval following transcranial stimulation of the inferior but not the superior parietal lobule. *Neuropsychologia, 51,* 900-906.

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain-regions associated with acquisition and retrieval of verbal episodic memory. *Nature, 368,* 633-635.

Smirni, D., Turriziani, P., Mangano, G. R., Cipolotti, L, & Oliveri, M. (2015). Modulating memory performance in healthy subjects with transcranial direct current stimulation over the right dorsolateral prefrontal cortex. *PLoS One, 10,* e0144838.

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

Susthana, N., Haneef, Z., Stern, J., Mukamel, R., Behnke, E., Knowlton, B., & Fried, I. (2012). Memory enhancement and deep-brain stimulation of the entorhinal area. *New England Journal of Medicine, 6,* 502-510.

Thompson-Cannino, J., Cotton, R., & Torneo, E. (2009*). Picking Cotton: Our memoir of injustice and redemption.* New York: St. Martin's.

Thomson, D. M. (1988). Context and false recognition. In G. M. Davies & D. M. Thompson (Eds.), *Memory in context: Context in memory* (pp. 285-304). Chichester, England: Wiley.

Tremblay, S., Lepage, J., Latulipe-Loiselle, A. Fregni, F., Pascual-Leone, A., & Théoret, H. (2014). The uncertain outcome of prefrontal tDCS. *Brain Stimulation, 7,* 773-783.

Troster, A. I., Wilkinson, S. B., Fields, J. A., Miyawaki, K., & Koller, W. C. (1998). Chronic electrical stimulation of the left ventrointermediate (vim) thalamic nucleus for the treatment of pharmacotherapy-resistant Parkinson's disease: A differential impact on access to semantic and episodic memory? *Brain and Cognition, 38,* 125-149.

Turriziani, P., Smirni, D., Zappalà, G., Mangano, G. R., Oliveri, M., & Cipolotti, L. (2012). Enhancing memory performance with rTMS in healthy subjects and individuals with Mild Cognitive Impairment: the role of the right dorsolateral prefrontal cortex. *Frontiers in Human Neuroscience, 6,* 62.

Turriziani, P., Oliveri, M., Salerno, S., Costanzo, F., Koch, G., Caltagirone, C., & Carlesimo, G. A. (2008). Recognition memory and prefrontal cortex: dissociating recollection and familiarity processes using rTMS. *Behavioral Neurology, 19,* 23-27.

Uncapher, M. R., Boyd-Meredith, J. T., Chow, T. E., Rissman, J., & Wagner, A. D. (2015). Goal-directed modulation of neural memory patterns: Implications for fMRI-based memory detection. *Journal of Neuroscience, 35,* 8351-8545.

Velanova, K., Jacoby, L. L., Wheeler, M. E., McAvoy, M. P., Petersen, S. E., & Buckner, L. (2003). Functional-anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *Journal of Neuroscience, 23,* 8460–8470.

Verfaellie, M., Rapcsak, S. Z., Keane, M. M, & Alexander, M. P. (2004). Elevated false recognition in patients with frontal lobe damage is neither a general nor a unitary phenomenon. *Neuropsychology, 18,* 94-103.

Vilberg, K. L. & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping, 30,* 1490-1501.

Wagner, A. D., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Prefrontal cortex and recognition memory: Functional-MRI evidence for context-dependent retrieval processes. *Brain, 121,* 1985-2002.

Wang, J. X., Rogers, L. M, Gross, E. Z., Rylas, A. J., Dokucu, M. E., Brandstatt, K. L., Hermiller, M. S., & Voss, J. L. (2014). Targeted enhancement of cortical-hippocampal brain networks and associative memory. *Science, 345,* 1054-1057.

Wang, J. X. & Voss, J. L. (2015). Long-lasting enhancements of memory and hippocampal-cortical functional connectivity following multiple-day targeted noninvasive stimulation. *Hippocampus, 25,* 877-883.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form.* Ames: The University of Iowa.

Wells, G. L. & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295.

Werner, N. S., Kuhnel, S., & Markowitsch, H. J. (2013). The neuroscience of face processing and identification in eyewitnesses and offenders. *Frontiers in Behavioral Neuroscience, 7,* 189.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. L., & Struss, D. T. (2002). Effects of time of day on age differences in working memory. *Journal of Gerontology, Series B., 57,* 3-10.

Wixted, J. T & Mickes, L. (2015). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition, 4,* 318-323.

Wheeler, M. E., & Buckner, R. L. (2003). Functional dissociation among components of remembering: Control, perceived oldness, and content. *Journal of Neuroscience, 23,* 3869–3880.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Zachle, T., Sandmann, P., Thorne, J. D., Janacke, L., & Herrmann, C. S. (2011). Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: Combined behavioral and electrophysiological evidence. *BMC Neuroscience, 12*.

Ziemann, U., Paulus, W., Nitsche, M. A., Pascual-Leone, A., Byblow, W. D., Beradelli, A., Siebner, H. R., Classen, J., Cohen, L. G., & Rothwell, J. C. (2008). Consensus: Motor cortex plasticity protocols. *Brain Stimulation, 1,* 164-82.

Zwissler, B., Sperber, C., Aigeldinger, S., Schindler, S., Kissler, J., & Plewnia, C. (2014). Shaping memory accuracy by left prefrontal transcranial direct current stimulation. *Journal of Neuroscience, 34,* 4022-4026.

APPENDIX A

Pilot Experiment: Forced Choice Task

We conducted the Pilot Experiment to more explicitly test the difficulty hypothesis in a well-controlled paradigm that manipulates task difficulty within one test. The paradigm used in this experiment was a two-alternative forced choice (2AFC) test in which participants studied lists of words in different font colors (red and green) and were later asked to identify which of two words was seen in one of the font colors. To manipulate difficulty, items were studied either one or three times. Test items on the 2AFC test were matched on difficulty such that each test trial contained a red and green word presented once or a red and green word presented three times. This particular task has been well-calibrated in previous studies in our lab (Sarfan & Gallo, unpublished), and thus should get around the previous projects' ceiling effects (Red or Green Word Test, Experiments 2 and 3, respectively) and floor effects (Red Word Test, Experiment 3). If tDCS on dlPFC is selective to the most challenging stimuli, then we would expect to see the largest stimulation benefits on the words studied once.

All participants were run in the morning because of the time of day findings in the first three experiments. We also had participants make confidence judgments at the end of each test trial, allowing us to explore the influence that tDCS of prefrontal cortex may have on local metacognition (Hertzog & Dunosky, 2011) in younger adults. Only two studies to the author's knowledge looked at trial-by-trial metacognitive ratings after applying stimulation to dlPFC. Chua and Ahmed (2016) found that HD-tDCS over left dlPFC led to improvements in the accuracy of judgments of learning about upcoming questions probing semantic memory. Bona and Silvanto (2014) and found that anodal stimulation led to a general reduction in confidence ratings but did not affect trial accuracy on a visual short term memory task. Finally, we included

a measure of working memory to see if we could parametrically predict the "benefit" provided by tDCS, although we did not find evidence for this parametric relationship.

We ran 14 participants in the left dlPFC stimulation condition and 16 participants in the sham condition. Consistent with the previous literature, left dlPFC stimulation led to small (though not statistically significant) improvements in working memory task performance. However, the groups did not differ in performance on the 2AFC memory test for items presented either once or three times. Furthermore, they did not differ in their post-trial confidence ratings on the 2AFC task for either correct or incorrect trials, suggesting no enhancement or detraction in local metacognition as a result of dlPFC stimulation.

This inability to conceptually replicate the findings of the previous studies highlights the critical importance of the cues available during retrieval. In Experiments 1-3, participants were only given one item to evaluate at a time during testing, during which time they had to compare the memorial evidence they had with some sort of criterion to determine whether to endorse a given item. In contrast, the 2AFC task allowed participants to compare the levels of memorial evidence they had for either item (or lack thereof) to each other. This suggests left dlPFC is not critical for these types of comparisons. Given that these initial 2AFC subjects did not show promising tDCS results, we decided to stop testing in this new paradigm and to instead go back to the single trial response format to determine if the morning results from the first three studies could be replicated in Experiment 4.