9-2018

The Relationship Between Cognitive and Neural Bases of Metamemory Judgments

Alexandra M. Gaynor

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Biological Psychology Commons, Cognitive Neuroscience Commons, and the Cognitive Psychology Commons

Recommended Citation


https://academicworks.cuny.edu/gc_etds/2798

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE RELATIONSHIP BETWEEN COGNITIVE AND NEURAL BASES OF METAMEMORY JUDGMENTS

by

ALEXANDRA MARGARET GAYNOR

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for

the degree of Doctor of Philosophy, The City University of New York

2018
The Relationship between Cognitive and Neural Bases of Metamemory Judgments

by

Alexandra Margaret Gaynor

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Date

Date

Elizabeth F. Chua, PhD
Chair of Examining Committee

Richard Bodnar, PhD
Executive Officer

Supervisory Committee:

Jennifer Mangels, PhD
Timothy Ellmore, PhD
Laura Rabin, PhD
Eric Leshikar, PhD

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

The Relationship between Cognitive and Neural Bases of Metamemory Judgments

by

Alexandra M. Gaynor

Advisor: Dr. Elizabeth F. Chua

Metamemory monitoring, the process of making subjective assessments of the status of one’s own memory, is crucial to guiding behavior and effective learning. Past cognitive research has shown that subjective confidence judgments are inferential in nature, and based on cues available at the time of the judgment. When confidence is based on cues that are related to objective memory performance, metamemory accuracy is high. However, past studies have shown that metamemory monitoring tends to be inaccurate because individuals base their confidence on information that is not predictive of memory success, such as the fluency with which items were encoded during study, or invalid information about task difficulty from external sources. Brain research has lagged behind cognitive research in establishing the neural bases of metamemory monitoring, but there is evidence that the prefrontal cortex contributes to Judgments of Learning (JOLs), subjective confidence judgments made during encoding about the ability to retrieve information at later test. Retrospective Confidence Judgments (RCJs), which are made at retrieval and reflect one’s confidence in his or her response, have been associated with activity in the prefrontal, parietal, and temporal lobes. However, there is evidence that the roles of each of these regions in metamemory monitoring may vary with the information on which confidence is based (e.g., fluency of encoding or information about task difficulty), and here we present two experiments which tested how the neural mechanisms underlying JOLs and RCJs interact with the cognitive bases of the judgments. Experiment 1 (Chapter 2) used transcranial direct current stimulation (tDCS) to directly manipulate activity in the dorsolateral prefrontal cortex (DLPFC) and anterior prefrontal cortex (aPFC) during a JOL task in which subjects studied words that varied in their fluency at encoding. Results showed that DLPFC stimulation impaired encoding for both fluent and disfluent conditions, and aPFC stimulation enhanced JOL accuracy for disfluent encoding conditions. Conversely, DLPFC and aPFC stimulation decreased JOL accuracy for fluently encoded conditions, suggesting the causal roles of the
aPFC and DLPFC in JOL accuracy vary depending on whether JOLs are based on the cue of fluency or disfluency.

Experiment 2 (Chapter 3) used fMRI to explore the neural correlates of RCJs and how confidence-related activity varies when RCJs are based on internal memory cues as compared to external cues about question difficulty during a semantic recognition task. Results showed that as compared to recognition, the process of making RCJs engaged regions consistent with those identified using episodic memory tasks. As compared to cued trials, uncued trials produced activity in lateral and medial parietal, and inferior temporal regions, which may reflect self-referential processing or uncertainty. Activity in the parietal lobe was greater for invalid than valid cues during hard questions, while occipital and subcortical regions showed greater activity during valid than invalid cueing for easy questions, suggesting confidence-related activity varies with both cue validity and actual question difficulty. Lastly, activity in bilateral medial temporal lobes covaried with the degree to which subjects incorporated invalid cues into their confidence judgments for correctly answered hard questions, which we speculate may reflect integration of multiple sources of information during RCJs. Taken together, these experiments suggest that the roles of brain regions involved in metamemory processes differ based on the information on which individuals base their confidence assessments. This work provides novel contributions to our understanding of the functional organization of the prefrontal cortex, as well as how multiple regions within the prefrontal, parietal, and temporal cortices contribute to recognition confidence based on external sources of information. Furthermore, our results inform the understanding of structure-function relationships in metamemory, which is critical to developing appropriate interventions to treat clinical populations with metacognitive deficits, and suggest non-invasive brain stimulation may be an effective method by which to alter metamemory performance.
Table of Contents

Chapter 1: General Introduction ........................................................................... 1

Chapter 2: Experiment 1: Transcranial direct current stimulation to the prefrontal cortex alters encoding and judgments of learning based on different sources of information ...................... 27
  Introduction ........................................................................................................ 29
  Methods ........................................................................................................... 34
  Results ............................................................................................................. 39
  Discussion ....................................................................................................... 45

Chapter 3: Experiment 2: fMRI activity associated with confidence in semantic recognition varies based on presence and validity of external cues ................................................................. 52
  Introduction ..................................................................................................... 54
  Methods .......................................................................................................... 60
  Results ........................................................................................................... 64
  Discussion ....................................................................................................... 69

Chapter 4: General Discussion ............................................................................ 78

Appendix A ........................................................................................................ 110

References ......................................................................................................... 111
List of Figures and Tables

Figures.................................................................................................................................98
  Figure 1...............................................................................................................................98
  Figure 2...............................................................................................................................98
  Figure 3...............................................................................................................................99
  Figure 4...............................................................................................................................99
  Figure 5.............................................................................................................................100
  Figure 6.............................................................................................................................101
  Figure 7.............................................................................................................................102
  Figure 8.............................................................................................................................103
  Figure 9.............................................................................................................................104
  Figure 10...........................................................................................................................105
  Figure 11...........................................................................................................................106

Tables...................................................................................................................................107
  Table 1...............................................................................................................................107
  Table 2................................................................................................................................108
Acknowledgments

Firstly, I would like to thank my advisor, Elizabeth Chua, who has been an exceptionally supportive and inspiring mentor. Thank you for your patience, perspective, and generosity, for always challenging me while still reminding me that I know more than I think I do, and for showing me the kind of mentor I hope to become.

I would also like to thank the members of my dissertation committee, Drs. Mangels, Rabin, Leshikar, and Ellmore, for their invaluable feedback, insight, and support throughout this process.

To the current and past members of the Memory and Metacognition and Human Cognition and Performance labs: I am so grateful for all the guidance, laughter, advice, and coffee breaks we’ve shared over the past five years. I’d also like to thank the members of my Behavioral and Cognitive Neuroscience cohort. I am honored to have completed this journey in the company of such a brilliant, creative, driven group of individuals, and am proud to call you not just colleagues, but lifelong friends.

I’d like to thank my family and friends, whose tremendous love and support over these years have made this dissertation possible. Margaret, words can’t begin to express the depth of my gratitude for your endless generosity and encouragement. Thank you for being a constant source of strength, humor, and love. To Katya, thank you for always challenging me to consider things from a different perspective, and inspiring me through your brave and ceaseless pursuit of your passions. Steven, your unwavering faith in me has been an invaluable source of confidence; thank you for always being my biggest cheerleader.

Finally, this dissertation is dedicated to my father, who always taught me not only to question the answers, but to question the questions themselves.

*All that we are arises with our thoughts. With our thoughts we make the world.*
Chapter 1
General Introduction

Unawareness of memory failures has been demonstrated not only in healthy individuals but also in populations with Alzheimer’s disease, schizophrenia, depression, and various learning disorders (Pannu & Kaszniak, 2005; Souchay, Isingrini, Pillon, & Gil, 2003). This unawareness can be viewed as a difficulty with metamemory, which has been broadly defined as knowledge about one’s own memory (Nelson & Narens, 1990). Metamemory is thought to guide subsequent behaviors, and may be especially useful for those experiencing difficulties in memory because accurate metamemory allows for using compensation strategies and seeking professional help that may result in earlier diagnosis and greater rehabilitation (Metcalfe, 2002; Souchay et al., 2003). However, more basic research is needed to understand the cognitive and neural bases of metamemory processes in healthy individuals, before we begin to understand how metamemory is affected in memory disorders.

One common cognitive framework of metamemory posits that cognitive processes are divided between an object level, which corresponds to memory processes, and a meta level, which corresponds to metamemory (Nelson & Narens, 1990). These levels are interrelated and information is exchanged through monitoring and control processes. Object-level information, such as the content of the memory, is monitored for accuracy and relevance based on the current goals, and the results of that monitoring inform the meta level, which contains a mental simulation of the object level. The object level can then, in turn, be controlled through behavior (Cavanaugh & Perlmutter, 1982). The monitoring stage of this process involves subjectively assessing one’s own memory abilities and the content of the object level (Nelson & Narens, 1990). On the other hand, metacognitive control involves the decisions and behaviors that are based on the meta-level representations that arise from the results of monitoring. For example, when a student is studying for a test, he or she must make a subjective judgment about how well he or she knows the information she is studying (i.e., a monitoring process), and then determine based on that judgment whether, and how much, more study is needed in order to recall the information at test (i.e., a control process). Metamemory monitoring judgments have been found to guide study time (Metcalfe, 2002; Nelson, Dunlosky, Graf, & Narens, 1994; Son & Metcalfe, 2000; Thiede, 1999) and also to influence error correction, in that feedback on inaccurate monitoring judgments results in those
inaccurately judged items being more likely to be corrected (Butterfield & Mangels, 2003). Thus, the ability to produce accurate metamemory judgments during monitoring is crucial to implementing effective behavioral control. Identifying the mechanisms underlying accurate metamemory monitoring is crucial to understanding how monitoring influences learning and memory in both healthy individuals and those with deficits.

Metamemory monitoring occurs at multiple time points along the learning and memory timescale, and further work is needed to understand monitoring processes that occur prior to retrieval (i.e., prospective monitoring) and after retrieval (i.e., retrospective monitoring) (Nelson & Narens, 1990). Prospective monitoring, which occurs prior to retrieval, reflects an individual’s confidence in his/her later ability to successfully retrieve studied items. One of the most commonly studied prospective monitoring tasks is the Judgment of Learning (JOL), a judgment made after study about one’s ability to later remember a currently retrievable item (Nelson & Narens, 1990). Retrospective monitoring tasks require subjects to monitor their memory after retrieval; the most common of such tasks is the Retrospective Confidence Judgment (RCJ), in which individuals judge their confidence in the accuracy of a previously retrieved response (Nelson & Narens, 1990). The experiments that will comprise this dissertation focus on the cognitive and neural bases of JOLs made during an associative encoding task (Experiment 1), and RCJs made during a semantic retrieval task (Experiment 2), in order to elucidate the mechanisms underlying metamemory judgments made at different points during the learning process. Although there is substantial behavioral research suggesting both these types of judgments may be inferential in nature and based on cues available at encoding or retrieval (Busey, Tunnicliff, Loftus, & Loftus, 2000; Jia et al., 2016; Koriat, 1997; Koriat & Ma’ayan, 2005; Koriat, Nussinson, Bless, & Shaked, 2008; Yue, Castel, & Bjork, 2012) there is significantly less research examining the neural mechanisms associated with metamemory processes (Chua, Schacter, Rand-Giovannetti, & Sperling, 2006; Chua, Schacter, & Sperling, 2009a, 2009b; Do Lam et al., 2012; Henson, Rugg, Shallice, & Dolan, 2000; Kao, Davis, & Gabrieli, 2005; Ryals, Rogers, Gross, Polnaszek, & Voss, 2016; Yokoyama et al., 2010). Moreover, there is a gap in knowledge about how the contributions of brain regions shown to be involved in metamemory monitoring may differ when subjective confidence is based on different cues at the time of the judgment. Given that accurate metamemory monitoring is crucial to effective learning and behavioral control
(Metcalfe, 2002; Nelson et al., 1994; Son & Metcalfe, 2000; Thiede, 1999; Thiede, Anderson, & Therriault, 2003), understanding the neural underpinnings of metamemory monitoring is an important topic of further investigation. The current studies aim to determine how regions of the brain known to be involved in metamemory processes may interact with the cognitive bases of JOLs and RCJs to contribute to metamemory monitoring processes in healthy adults.

The Cognitive Basis of Metamemory

Early theories regarding the cognitive basis of metamemory proposed that individuals have privileged access to their own memory traces, and that metamemory judgments are made based on the strength of this memory trace (Hart, 1965). However, a significant body of experimental research has demonstrated that some manipulations alter metamemory ratings but not memory accuracy, and vice versa, posing a challenge for the hypothesis that subjective judgments are based on direct access to memory strength (Daniels, Toth, & Hertzog, 2009; Dunlosky & Nelson, 1994; Rabinowitz, Ackerman, Craik, & Hinchley, 1982; Shaughnessy, 1981). Instead, it has been widely accepted that metamemory judgments are inferential in nature and based on information available at the time of the judgment, which may include the strength of the memory trace, but may also include cues such as the fluency with which the stimulus is processed, or the conditions of learning (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Koriat, 1997). Because JOLs and RCJs are made at different points throughout the learning process, different information is available for individuals to use as the basis of their metamemory judgments. For instance, JOLs, which are made during encoding, can be based on cues related to ease of encoding or familiarity of the stimulus, whereas RCJs are likely based on cues related to retrieval processes, such as speed of retrieval or amount of information retrieved. When individuals base their inferential judgments on cues that influence memory performance, their judgments are likely to be accurate (i.e., the metamemory judgment and the memory judgment are congruent, such as a high confidence correct response). However, there are many cues available that appear to influence JOLs and RCJs, but not actual memory success, such as fluency of encoding (Hertzog, Dunlosky, Robinson, & Kidder, 2003; Jia et al., 2016), and invalid information from external sources (Bradfield, Wells, & Olson, 2002; Jaeger, Lauris, Selmeczy, & Dobbins, 2012). When individuals base their inferential judgments on such cues that
are not indicative of memory success, these judgments are inaccurate (i.e., the metamemory judgment and the memory judgment are incongruent, such as a high confidence incorrect response). Thus, in order to understand what drives metamemory accuracy, it is important to understand what cues individuals use to make JOLs and RCJs and how these cues relate to memory accuracy.

Judgments of Learning

JOLs, prospective judgments made at study about the ability to remember given items at a later test, are one of the most commonly studied types of metamemory monitoring (Dunlosky & Nelson, 1992; Koriat, 1997; Nelson & Dunlosky, 1991; Nelson & Narens, 1990), and the level of JOL rating given is often predictive of strategic control of memory processes, suggesting individuals use JOLs to guide their learning (Metcalfe, 2002; Nelson et al., 1994; Son & Metcalfe, 2000; Thiede, 1999). JOL accuracy, which is a measure of how well the judgment at study predicts actual subsequent performance at test, has been correlated with better memory performance (Nelson et al., 1994; Thiede et al., 2003), such that individuals with better metamemory also tend to have better memory. However, individuals with intact memory often demonstrate inaccurate JOLs (Dunlosky & Nelson, 1992), suggesting memory and metamemory processes are dissociable, and further research is necessary to investigate the mechanisms underlying accurate metamemory.

In an experimental setting, trial-by-trial JOLs are typically assessed by having participants study stimuli, most often words, but sometimes visuospatial or facial stimuli (Nelson, Narens, & Dunlosky, 2004; Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995; Watier & Collin, 2011), and then generate a JOL by rating their confidence in their ability to remember a given item at later test. Alternatively, subjects may be asked to give global JOLs, during which they estimate how many items total they think they will remember at test. Trial-by-trial JOLs are often given on a scale of 0-100% confidence (Kelemen, 2000; Scheck, Meeter, & Nelson, 2004) or a Likert-type point scale (King, Zechmeister, & Shaughnessy, 1980), but subjects may also be prompted to rank items by how well they think they will remember them (Leonesio & Nelson, 1990). JOLs can be given immediately following the presentation of each item, known as immediate JOLs, or following a given amount of time, usually lasting the duration of the remainder of the study set, referred to as delayed JOLs. After each item has been studied an given a
JOL, a memory test is given to test either recognition or recall of studied information, which enables analyses of metamemory accuracy (Dunlosky & Nelson, 1992; Nelson & Narens, 1990).

One early hypothesis regarding the cognitive basis of JOLs is that JOL ratings are based directly on encoding strength (Arbuckle & Cuddy, 1969; Daniels et al., 2009; King et al., 1980). This hypothesis echoes more general ‘direct access’ or ‘trace access’ theories of metamemory monitoring, which suggest that metamemory judgments are based on the strength of traces of the actual memory being judged (Arbuckle & Cuddy, 1969; Hart, 1965, 1967a, 1967b; Sikström & Jönsson, 2005). Indeed, there is evidence that stronger memories may lead to higher and more accurate JOLs. Daniels et al. (2009) had younger and older adults make immediate JOLs during study, and then rate the quality of their memory for each item using Recollect, Familiar, or No Memory judgments at test. They found that items recollected at test were correlated with higher JOL ratings at study as compared to familiar or not remembered items (Daniels et al., 2009). Such evidence of a positive correlation between memory performance and level of metacognitive confidence is consistent with the hypothesis that JOL ratings may be partially based on direct access to the strength of the memory being judged, or based on factors that influence subsequent memory accuracy, such as the amount of detail included when generating a mental image at study.

Despite the above evidence that subjective confidence may be somewhat based on memory strength, one important limitation of direct-access theories lies in the idea that if JOLs or other metamemory judgments are in fact based directly on memory abilities, experimental manipulations that affect JOLs should affect memory performance in similar ways, and vice versa. While this tends to be the case, both processes are not always sensitive to the same manipulations (Dunlosky & Nelson, 1994; Rabinowitz et al., 1982; Shaughnessy, 1981; Skavhaug, Wilding, & Donaldson, 2010). For instance, Shaughnessy (1981) conducted an experiment in which subjects studied paired-word associates using either interactive imagery or rote rehearsal strategies, and made an immediate JOL following each pair. Participants underestimated memory abilities using imagery as compared to rote rehearsal: memory performance was greater for pairs encoded using imagery, but there was no difference in JOLs between groups (Shaughnessy, 1981). These results oppose the direct access views that JOLs are based directly on memory traces, given that manipulations affecting memory performance failed to influence JOL
ratings.

An alternative to the direct access theory is that metamemory judgments are inferential, and based on various cues available at the time of the judgment (Koriat, 1997). These may include the conditions of learning, characteristics of the stimuli being learned, or the perceived strength of the memory trace (Koriat, 1997). One cue that has been shown to influence JOL ratings is encoding fluency, which is typically measured by the speed involved in executing a task (Begg et al., 1989; Hertzog et al., 2003). For example, when participants are asked to use imagery to connect paired items during encoding, fluency would relate to the speed with which the image is created, how many images are generated, or how salient the image is (Begg et al., 1989; Hertzog et al., 2003). Numerous studies have suggested that JOLs are positively correlated with speed or ease of processing (Begg et al., 1989; Benjamin, Bjork, & Schwartz, 1998; Hertzog et al., 2003). Hertzog et al. (2003) measured fluency using subjects’ latency in generating mental images to bind the two items in a word pair during encoding, and found a strong correlation between latency in generating an image and JOL ratings. However, fluency was not correlated with actual recall performance, indicating that individuals’ JOLs were based on a sense of ease of processing, but this did not actually contribute to JOL accuracy, as JOLs were inconsistent with probability of recall (Hertzog et al., 2003). Benjamin, Bjork and Schwartz (1998) also found that JOLs were positively correlated with response time during recall of trivia information, but slowly-produced answers were associated with more successful retrieval. Begg et al. (1989) found that the longer participants took to study items in a paired-associate task, the lower their JOLs, indicating that JOLs may have been driven by heuristics related to ease of encoding, such as the idea that items requiring more study time are less likely to be remembered (Begg et al., 1989). Similarly, Benjamin (2003) found that subjects predicted they would be more likely to recognize high-frequency as compared to low-frequency words, but actually had better recognition memory for low-frequency words (Benjamin, 2003). In a more recent study, Jia et al. (2016) found that manipulating perceptual fluency by altering font style did not alter the effect of word frequency on JOLs, and that subjects gave higher estimates for hypothetical participants’ ability to remember high- as compared to low-frequency words, suggesting that a combination of fluency and pre-existing beliefs about stimulus memorability may jointly contribute to inaccurate JOL ratings.
Although JOL ratings are positively correlated with processing fluency, there is a great deal of evidence that information more deeply processed at encoding is better remembered, making disfluency a better predictor of future memory success (Craik & Lockhart, 1972; Craik & Tulving, 1975). However, individuals tend to discount the memory benefits of ‘desirable difficulties’: those conditions which produce less fluent encoding but lead to better learning because they are more deeply processed (Bjork, 1994; Sungkhasettee, Friedman, & Castel, 2011). Perceptual disfluency is one such desirable difficulty that promotes deeper encoding but is generally discounted by subjects making metamemory judgments. Studies have shown that manipulations that promote disfluency, such as difficult-to-read fonts, produce better memory performance than easy-to-read fonts (Diemand-Yauman, Oppenheimer, & Vaughan, 2011). One study has suggested that in situations of limited visual information, subjects must generate the full stimulus themselves, producing a generation effect that benefits memory (Nairne, 1988). Although perceptually disfluent items are likely better remembered due to more elaborative processing during encoding, individuals’ metamemory judgments often do not account for this memory benefit. Sungkhasettee et al. (2011) showed that participants’ JOLs were similar for words presented upright and upside-down at study, but they later recalled more upside-down than upright words due to deeper processing of upside-down words during encoding.

Similarly, subjects disregard the benefit of deep encoding strategies when studying paired associates rehearsed under “maintenance” or “elaborative” instructions: they have better recall for items under elaborative rehearsal, but subjectively judge both strategies to be equally effective (Shaughnessy, 1981). Subjects’ JOLs are also insensitive to the memory benefits of using interactive imagery during encoding (Rabinowitz et al., 1982), and to the improved recall resulting from repeated presentations of stimuli as compared to single presentations (Koriat, 1997). This effect persisted even when participants were given feedback about the accuracy of their answers, suggesting a strong tendency to discount the beneficial effects of repeated stimulus presentation.

These findings are in line with theories of cue-utilization that distinguish between intrinsic and extrinsic cues as bases for JOLs. Intrinsic cues include characteristics inherent to the stimuli themselves, such as familiarity of the stimulus, while extrinsic cues include manipulations at encoding, such as repetition or duration of item presentation (Koriat, 1997). Koriat (1997) demonstrated that in a paired-
associates memory task, intrinsic cues had similar effects on recall and JOLs, but extrinsic factors did not influence JOLs as strongly as they did performance, suggesting the basis of JOLs, and the relationship between JOLs and encoding processes, may be mediated by the type of cue on which people based inferences about memorability (Koriat, 1997). Several other studies have provided evidence that JOLs are sensitive to intrinsic cues, but tend not to take into account extrinsic cues, even though the latter are often more predictive of memory success (Begg et al., 1989; Koriat, 1997; Koriat & Ma'ayan, 2005). JOLs that are based on cues that influence later performance will be more accurate than those based on information that is not predictive of memory success, and determining how individuals use these cues appropriately to support accurate metamemory predictions is critical to developing interventions that may promote effective metamemory control to enhance learning.

*Retrospective Confidence Judgments*

Unlike JOLs, which are prospective judgments made at or immediately following study, RCJs are retrospective judgments, which are made after retrieval and reflect the individuals’ confidence in the accuracy of their response (Nelson & Narens, 1990). Experimental tasks that measure RCJs may assess confidence following either recall or recognition. Recognition judgments may be one-step responses, in which participants are asked to give a recognition judgment that also incorporates their confidence in the response (for example, in an old/new recognition test, indicating ‘sure old’, ‘unsure old’, ‘unsure new’ or ‘sure new’) (e.g., Kim & Cabeza, 2007, 2009; Moritz, Gläscher, Sommer, Büchel, & Braus, 2006), or using a two-step process, in which participants give an RCJ immediately following the recognition response on confidence scale (e.g., Chua et al., 2006; Chua, Schacter, & Sperling, 2009a).

Behavioral research has suggested that similar to JOLs, RCJs are inferential in nature, and are based on information available at retrieval, in addition to cues at encoding (Busey et al., 2000; Koriat & Levy-Sadot, 1999; Koriat et al., 2008). There is some evidence that subjects monitor and assess the strengths of their stored memories at the time of retrieval, and base their confidence on the results of these assessments (Busey et al., 2000). Cues directly related to memory, such as vividness, are a common basis of RCJs, and often lead to relatively accurate confidence judgments because confidence and actual memory performance are related to the same information (Busey et al., 2000).
Although findings that RCJs are based on cues that directly influence memory strength is consistent with a direct access view of metamemory, RCJs are also often based on information that is nondiagnostic of actual memory performance. For instance, in an eyewitness memory paradigm, Lindsay et al. (1998) found that when individuals made recognition confidence judgments following identification of a target from a lineup, conditions that led to better memory performance (e.g., viewing a 3 minute video of the target from a variety of perspectives, as compared to a 10 second video of a brief close-up of the target) also led to increased confidence, suggesting accuracy and confidence were based, at least in part, on the same information. However, conditions that resulted in higher RCJs for correct recognition also resulted in higher confidence in false identifications, suggesting that participants tended be more confident overall under conditions they expected would lead to better memory, even when their responses were incorrect (Lindsay, Read, & Sharma, 1998). Thus, despite evidence that RCJs can be partially based on the strength of the memory being retrieved, findings of high confidence in false memories indicate that confidence and memory are at least partially dissociable and rely on somewhat different mechanisms. This is consistent with the theoretical framework of metamemory, in which the meta-level contains a dynamic representation of the memory itself, which is often imperfect due to its susceptibility to distortions, which are typically caused by failure to incorporate information that is present in the object-level, or incorporating properties that are not actually present (Nelson & Narens, 1990).

Evidence that individuals have high confidence in incorrect responses challenges early direct access views of metamemory, and more recently, a growing body of research has suggested that RCJs are likely inferential in nature and often based on information other than the strength of the memory, some of which may not be reflective of actual memory success. For instance, Loftus & Pickrell (1995) found that post-event suggestions can lead individuals to remember a wide variety of events that did not actually happen to them, with no awareness of the inaccuracy of these memories. In remember-know paradigms testing recognition of word stimuli, subjects made ‘remember’ false alarms to nonstudied items, even though a ‘remember’ judgment typically reflects a subjective sense of a stronger or more vivid memory trace than a ‘know’ responses (Norman & Schacter, 1997; Roediger & McDermott, 1995). Thus, it appears that recognition judgments are not based entirely on access to the strength of memory traces. Instead, such studies provide evidence that RCJs are in fact inferential and based on cues available at
retrieval, such as familiarity and fluency (Busey et al., 2000; Koriat et al., 2008). This theory is consistent with Koriat’s (1995) accessibility hypothesis, which posits that individuals search through information in memory at retrieval, and base their RCJs on whatever information they are able to retrieve (Busey et al., 2000; Koriat, 1995, 1997). This information may include the memory trace, but may also be comprised of other information such as familiarity of the cue or target, or the conditions of retrieval (Busey et al., 2000).

There is evidence that individuals rely on multiple sources of information when assessing their confidence, and Koriat et al. (2008) proposed that the information on which subjects base RCJs can be broadly divided into experience-based and information-based cues (Koriat & Levy-Sadot, 1999; Koriat et al., 2008). Information-based judgments are those that rely on pre-existing beliefs about one’s memory, competence in the domain being tested, or exam difficulty. They may also be based on the strength of evidence that is gathered in favor of the chosen response, relative to evidence in support of alternatives (Koriat et al., 2008). Experience-based judgments, in contrast, are not based on declarative content, but rather, rely on subjective feelings or cues that arise from the experience of learning, such as a sense of familiarity with the stimulus or the ease with which an answer was retrieved. Confidence judgments may be affected by both information-based and experience-based sources of information, reflecting both declarative information about the item being retrieved, and the subjective sense of ease with which it was retrieved (Koriat et al., 2008).

One experience-based cue that has been shown to influence RCJs is retrieval fluency (Shaw, 1996; Shaw & McClure, 1996). Shaw & McClure (1996) repeatedly questioned participants about an event they witnessed over the course of several weeks following the event, and showed that subjects had greater confidence for those details about which they were repeatedly asked during post-event questioning, as compared to those items they were only asked about during the initial questioning. However, actual memory accuracy was not significantly different between conditions, indicating repeated retrieval during post-event questioning may lead to a greater sense of fluency, which artificially inflates memory confidence. Shaw (1996) also found that simply engaging in reflective thought about previous responses caused subjects to have higher later confidence ratings for both correct and incorrect responses, but had no impact on later memory accuracy, again supporting the idea that individuals base RCJs on a sense of retrieval fluency despite this ease of retrieval not being associated with improved
memory. Similarly, Kelley & Lindsay (1993) used a priming procedure to demonstrate that prior exposure to both correct and related, but incorrect, answers on a general knowledge test influenced subjects’ response times and confidence in giving primed answers. Repeated presentation of a scene has also been shown to increase confidence ratings for scene-face associative recognition, especially for incorrect responses, suggesting that individuals base RCJs on familiarity with the cue, even though this may not reflect accurate retrieval of the cue-target pair (Chua, Hannula, & Ranganath, 2012).

Support for the hypothesis that RCJs are also based on information-based cues has come from research showing that confidence is sensitive to consensuality, i.e. the proportion of people who endorse a given response, regardless of whether that response is correct (Koriat, 2008b). The effect of consensuality on confidence suggests that individuals rely on informational cues to make memory judgments, increasing their confidence for questions to which they believe most people should know the answer. In one experiment, subjects studied English words and their translation in a foreign language, including word-pairs for which participants are likely to agree on the wrong translation (Koriat, 2008b). Results showed that confidence ratings correlated with the consensuality of the translation rather than its correctness, as subjects had increased confidence for the correct answer when the majority of participants agreed on the correct answer, but also had inflated confidence for the incorrect answer when it was a response endorsed by most participants (Koriat, 2008b). These findings suggest confidence is sometimes based on information-based cues. Because responses endorsed by others are correct more often than not (Koriat, 1976), items for which there is a consensus on the correct answer likely activate more correct than incorrect information about the target. However, items for which there is a consensus on the incorrect answer likely also activate many incorrect partial clues about the target, suggesting that subjects use an information-based process by deliberately weighing evidence in favor of one response over another when making their confidence judgments, regardless of whether that evidence is correct.

Confidence in incorrect responses may rely on experienced-based cues as well, including cues that influence confidence in correct answers, such as ease of retrieval (Kelley & Lindsay, 1993). Indeed, there is some evidence that experience-based processes contribute to the correlation between confidence and consensuality: response times are shorter for consensual responses, regardless of their correctness. This suggests subjects may use response latency as an experience-based cue on which to
base their confidence (Koriat et al., 2008), and that confidence judgments likely rely on a combination of both information-based and experienced-based cues.

Other studies have provided further evidence that subjects use a combination of information- and experience-based information when making RCJs (Jaeger, Lauris, et al., 2012; Koriat et al., 2008). One experiment demonstrated that when subjects were asked to spontaneously list reasons in support of an answer, confidence increased with the number of reasons listed, presumably because of both increased supportive evidence and greater ease of retrieval (Koriat et al., 2008). However, when researchers imposed the number of reasons to be listed, confidence did not increase with number of reasons subjects were required to list. This suggests that the increased effort subjects experienced with a greater number of reasons in the forced-report condition counteracted the effect of the strength of supportive evidence on confidence. Thus, it appears that when information-based cues (such as strength of supporting evidence, reflected by number of reasons listed) and experience-based cues (such as ease of retrieval) conflicted, confidence judgments were made on the basis on the experience of retrieval (Koriat et al., 2008).

Although in some cases individuals appear to base their confidence on experience-based cues when they conflict with information-based cues, there is also evidence that people rely more heavily on external information when unable to rely on internal experience-based signals to make RCJs (Jaeger, Lauris, et al., 2012). In one episodic memory task, subjects were given cues from reliable (75% valid) and unreliable (50% valid) sources about whether an item was ‘likely old’ or ‘likely new’ in an old/new recognition test (Jaeger, Lauris, et al., 2012). For low confidence trials, when participants could not rely on their own internal cues, they incorporated the recommendations of both reliable and unreliable sources, and continued to rely on cues from unreliable sources even when those sources were only correct on 25% of trials (Jaeger, Lauris, et al., 2012). Subjects have also been shown to base their estimates of others’ ability to solve anagrams on their own ease of solving them, but when answers were provided for them, eliminating the subjective sense of fluency they derived from solving the problems themselves, subjects instead based their judgments on pre-existing theories and rules about solving such problems (Kelley & Jacoby, 1996).

The hypothesis that individuals rely more heavily on information-based cues when their internal experience-based cues are weaker is further supported by the finding that inaccurate eyewitnesses’
信心比提供确认反馈影响更大，准确的目击者在 lineup（Bradfield et al., 2002）中。这种后确认反馈效果很可能发生因为个体不参加在线监控他们的信心或其他变量，与回忆条件，如他们对嫌疑人的看法或速度的确认。由于这种弱记忆力的条件，回忆者更依赖反馈来做出信心判断。一致的与想法相符，主体的判断受到外部影响时，Wells & Bradfield (1999) 发现主体被要求私下思考他们的确定性之前，反馈，他们更少被影响的后事件反馈，可能因为更强烈的内部的关于他们检索反应准确性影响的内部信号，导致更少依赖外部因素，当做出RCJ。综合一起，研究发现，主体如何纳入外部信息到他们的信心判断，表明个体更可能基于基于信息的内部线索，如那些关系到记忆和实际记忆痕迹的强度，是不可靠的。

Metamemory Accuracy Measures

识别哪些线索主体用来做出元记忆判断，允许我们识别在准确元记忆中的认知机制，以及那些影响主效应和实际记忆性能的变量。通常，元记忆准确性通过测量主效应和实际记忆性能的关联程度来评估（Benjamin & Diaz, 2008）。在实验元记忆任务中，元记忆准确性是指主观的判断与实际的回忆性能在测试，即，正确的响应被确认与高信心和错误的响应被确认与低信心（Maniscalco & Lau, 2012）。

元记忆准确性可以分为那些测量绝对值与相对值。绝对准确性反映了平均评级值和平均性能的关系，而相对准确性反映了条件下内相关性。与记忆性能和元认知预测（Benjamin & Diaz, 2008）。校准曲线是测量绝对准确性的常用方法，其中平均性能被映射出来。
(Benjamin & Diaz, 2008). These provide a measure of the degree to which mean ratings correlate with mean performance, i.e. and individual with perfect calibration between memory and metamemory would give a mean confidence of 80% to items they remembered with 80% accuracy (Benjamin & Diaz, 2008).

Relative accuracy measures are often preferred to absolute accuracy measures, such as calibration curves, because they are less influenced by idiosyncratic ways in which subjects use the rating scale (Scheck et al., 2004). Instead, relative accuracy measures provide a measure of correspondence between metamemory judgments and memory performance that is based on confidence of one item relative to another (Nelson & Narens, 1990). Historically, the most commonly used statistical tool for analyzing relative metamemory accuracy is the Goodman-Kruskal Gamma Coefficient, γ, which measures the association between confidence judgments and memory performance based on the difference between concordant and discordant pairs (Benjamin & Diaz, 2008; Nelson & Narens, 1990), calculated by the formula \((\text{Concordances} - \text{Discordances})/(\text{Concordances} + \text{Discordances})\) (Spellman, Bloomfield, & Bjork, 2008). The gamma coefficient is useful in that its expected value is constant across any changes in a person’s confidence threshold, and requires no assumptions be met regarding distribution of ratings (Nelson & Narens, 1990). However, the gamma coefficient has been criticized as a suboptimal measure of relative accuracy due to the fact that it treats data ordinally, and does not take into account the magnitude of judgments: for instance, if a subject were to assign two items JOLs of 5% and 95% confidence, gamma would treat this equivalent to a case in which a subject rated two items with JOLs of 49% and 50% (Benjamin & Diaz, 2008). Interval-properties, which allow for the measurement of the relative degree of difference between items, may prove to be a better measure of metamemory monitoring accuracy.

To address the need for interval-level analysis, a signal-detection based measure called \(d_a\) has sometimes been used in addition to, or instead of, gamma coefficients in studies of metamemory accuracy (Benjamin & Diaz, 2008; Masson & Rotello, 2009; Toth, Daniels, & Solinger, 2011). In metacognitive signal detection, a metacognitive hit, such as giving a high confidence rating to a successfully remembered item, is plotted against the probability of a metacognitive false-alarm, such as a high-confidence item that is not successfully remembered. In a task with a confidence rating scale from 1-10, each confidence level would be treated as a response criterion, which distinguishes high from low
confidence, such that a liberal criterion would assign low confidence to a response of 1 and high confidence to responses of 2-10, and a conservative criterion would assign low confidence to responses of 1-9 and high confidence to ratings of 10. Each possible split of the response data produces hit and false alarm rates that are plotted against one another. $D_a$ represents the distance between this isosensitivity plot and the line representing chance performance (Benjamin & Diaz, 2008), and is calculated by the following formula: $d_a = \sqrt{2y_0}\sqrt{1+m^2}$ in which $y_0$ represents the y-intercept and $m$ represents the slope of the isosensitivity function.

Another calculation of metamemory accuracy that is closely related to $d_a$ uses receiver operating characteristic (ROC) analyses (Fleming & Lau, 2014). Similarly to $d_a$, this method involves plotting metacognitive ‘hits’ (high-confidence correct) against ‘false alarms’ (high-confidence misses), and metacognitive sensitivity is reflected by the area under the ROC curve, typically abbreviated as AUROC2. If the AUROC2 is 0.5, metacognitive accuracy is at chance, with values above 0.5 indicating more accurate metacognitive monitoring (Fleming & Lau, 2014). Both $d_a$ and AUROC2 are useful in that they do not require that the underlying distributions of memory responses and metamemory judgments be equal in variance, and allow for interval-level analyses with equal intervals across the scale range (Benjamin & Diaz, 2008).

Another model-based approach to measuring metamemory accuracy, known as meta-$d'$, allows researchers to dissociate task performance from metacognitive sensitivity using signal detection analyses (Maniscalco & Lau, 2014). Because performance on the memory task itself can influence metamemory sensitivity, meta-$d'$ accounts for this influence by comparing metamemory sensitivity to primary task sensitivity ($d'$). Meta-$d'$ and $d'$ are measured in the same signal-to-noise ratio units, making it possible to directly compare the sensory evidence available for decision-making ($d'$) to the sensory evidence available for metacognition (meta-$d'$) (Fleming & Lau, 2014). Thus, an ideal subject who uses the maximum possible metacognitive sensitivity to make confidence judgments would have equal meta-$d'$ and $d'$ values. This approach is particularly useful because it is a measure of metamemory efficiency, i.e. metamemory performance relative to underlying task performance, and thus can account for differences in memory performance (Fleming & Lau, 2014). This dissertation reports $d_a$ measures of metamemory accuracy, which will allow us to use improved ways of indexing metamemory accuracy and also compare
our results to other metamemory literature using this measure.

The Neural Basis of Metamemory Processes

Although a significant amount of research has been conducted on the cognitive bases of metamemory processes, relatively less is known about the neural bases of metamemory accuracy. Most of the neuroimaging and lesion studies addressing metamemory function have focused on regions that modulate by the subjective level of confidence ratings (Chua et al., 2006, 2009a; Do Lam et al., 2012; Kao et al., 2005), and only a few studies have identified brain regions associated with metamemory accuracy (Kao et al., 2005; Yokoyama et al., 2010). One potential reason for the lack of findings reported about metamemory accuracy is that different components of metamemory accuracy, such as knowing when you know (i.e., high confidence in hits) and knowing when you don’t know (i.e., low confidence in misses) may rely on different neural and cognitive mechanisms (Luo, Kazuhisa, & Luo, 2003). Although there has been work associated with metamemory accuracy and the subjective levels of confidence, based on the cognitive research there is a need to further our knowledge about the brain bases of metamemory processes relates metamemory judgments that are based on different cues.

Neuroimaging and lesion studies have implicated regions of the prefrontal cortex in JOL processes (Do Lam et al., 2012; Kao et al., 2005; Vilkki et al., 1999; Yokoyama et al., 2010), and prefrontal and parietal cortices in RCJs (Chiou, Carlson, Arnett, Cosentino, & Hillary, 2011; Davidson et al., 2008; Pannu & Kaszniaik, 2005; Simons, Peers, Mazuz, Berryhill, & Olson, 2010), but there is some evidence that this activity may vary with the type of judgment and the information on which the judgment is based (Kim & Cabeza, 2007; Ryals, Rogers, Gross, Polnaszek, & Voss, 2016). Cognitive research has shown that metamemory judgments may be based on various sources of information, and those made on information that affects memory performance, such as depth of processing or internal signals regarding the strength of a memory trace, are more accurate than judgments made on the basis of information that is not diagnostic of memory accuracy, such as encoding fluency or external information (Koriat, 1997; Koriat et al., 2008). Investigating how the recruitment of different brain regions may vary with the cognitive bases of metamemory judgments is crucial to our understanding of accurate memory awareness, but remains a relatively unexplored avenue of research. Experiments 1 & 2 examine how the
brain bases of JOLs and RCJs, respectively, may vary based on the cognitive bases of the metacognitive judgment.

Judgments of Learning recruit the prefrontal cortex

Past research has typically focused on the cognitive bases of metamemory processes (Jaeger, Lauris, et al., 2012; Kelley & Lindsay, 1993; Koriat, 2008b; Koriat & Levy-Sadot, 1999; Koriat et al., 2008; Shaw & McClure, 1996), but only a handful of studies have investigated the neural bases of metamemory monitoring (Do Lam et al., 2012; Kao et al., 2005; Ryals et al., 2016; Vilkki et al., 1999; Yokoyama et al., 2010), and further studies are needed to investigate the neural underpinnings of JOLs and how they relate to the cognitive bases of these judgments. Several neuroimaging and lesion studies have provided evidence that the PFC contributes to JOL ratings, encoding, and accuracy (Do Lam et al., 2012; Kao et al., 2005; Vilkki, Servo, & Surma-aho, 1998; Vilkki et al., 1999). For example, patients with both right and left prefrontal lesions are less accurate than healthy controls and posterior lesion patients in predicting future retrieval success (Vilkki et al., 1998), and those with right frontal lesions showed impaired JOL accuracy in a face-location association task (Vilkki et al., 1999), suggesting a causal role of the PFC in metamemory accuracy. Neuroimaging experiments have provided further evidence of how specific prefrontal subregions contribute to JOL processes. Specifically, areas in the anterior prefrontal cortex (aPFC), comprised of Brodmann areas 10 and 11, have been implicated in JOL processes (Fig. 1) (Do Lam et al., 2012; Kao et al., 2005). Two fMRI studies have demonstrated increased medial prefrontal (BA10) and orbitofrontal (BA11) activity with higher JOLs, suggesting these regions may track the magnitude of the subjective level of confidence (Do Lam et al., 2012; Kao et al., 2005). Kao et al. (2005) also showed that activity in more lateral and posterior regions of the PFC (BA44/6) was associated with both JOL ratings and encoding, signifying common underlying mechanisms between JOL and encoding processes. Turning to metamemory accuracy, individual differences in JOL accuracy were correlated with ventromedial prefrontal activity (BA11), suggesting that in addition to being involved in tracking the magnitude of JOL ratings, more anterior regions of the PFC play a role in JOL accuracy. This distinction between functional roles of the anterior and posterior PFC in metamemory is consistent with studies showing that various functions of cognitive control are arranged hierarchically along the rostro-caudal axis
of the PFC in a gradient of increasing complexity (Badre, 2008; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Koechlin, Ody, & Kouneiher, 2003).

Past research on the functional organization of the PFC has suggested that posterior regions of the PFC track information about stimulus properties and relay this information to anterior regions, which are responsible for forming more abstract representations and relationships (Badre, 2008; Badre & D’Esposito, 2009; Koechlin et al., 2003). Indeed, there is evidence that the DLPFC is activated when a task requires monitoring and manipulation of externally-generated information, such as information about the task stimuli, while the aPFC is involved in monitoring information that is internally-generated, such as overarching task goals or self-referential assessments of performance (Christoff & Gabrieli, 2000). Thus, if metamemory functions are similarly arranged in a hierarchical fashion within the PFC, it is possible that the DLPFC modulates with the level of JOL ratings because posterior regions track stimulus-level information that influences confidence, while the aPFC is associated with JOL accuracy because it is responsible for monitoring the relationship between JOLs and encoding in the interest of ongoing task goals. In addition to neuroimaging evidence that supports a correlation between the aPFC and metamemory accuracy, one study has also demonstrated a causal role of the aPFC in JOL accuracy. Ryals et al. (2016) applied theta-burst stimulation, a form of non-invasive brain stimulation, to areas of the PFC during a JOL task, and showed that compared to DLPFC and sham stimulation, aPFC stimulation decreased overall JOL ratings for subsequent misses, suggesting the aPFC plays a causal role in JOL accuracy, and that stimulation made subjects more aware of failures at encoding that would result in poor performance at test. This raises the hypothesis that perhaps rather than integrating all lower-level signals from posterior regions, the aPFC supports JOL accuracy by increasing sensitivity to information at encoding that is predictive of memory performance (Ryals et al., 2016).

There is a clear need to examine how the roles of PFC subregions involved in JOL processes may vary depending on whether JOLs are based on cues that are either diagnostic or nondiagnostic of future memory performance. Elucidating the specific roles of PFC regions in JOL accuracy, and how they may change with the cognitive bases of the judgments, would provide insight into how the brain supports accurate metamemory awareness, as well as the general functional organization of the PFC. Experiment 1 manipulates the basis of JOLs to test whether the aPFC is correlated with JOL accuracy because it
integrates information from posterior regions (Badre, 2008), or because it increases sensitivity to diagnostic cues over nondiagnostic cues (Ryals et al., 2016). Experiment 1 will use a non-invasive brain stimulation technique known as high-definition transcranial direct current stimulation (HD-tDCS), which allows us to directly manipulate brain activity to test hypotheses about the causal roles of PFC regions in JOLs, and also determine whether brain stimulation may serve as an effective way to enhance JOL accuracy.

Retrospective Confidence Judgments rely on prefrontal and parietal regions

Neuropsychological experiments have provided evidence that the prefrontal and parietal cortices are involved in RCJs (Chiou et al., 2011; Davidson et al., 2008; Kennedy, 2001, 2004; Simons et al., 2010). Although some studies have shown that patients with frontal traumatic brain injuries demonstrated intact RCJs following recall (Kennedy, 2001, 2004), others have demonstrated that RCJs were impaired in frontal lobe damaged patients in recognition tasks (Chiou et al., 2011; Pannu, Kaszniak, & Rapcsak, 2005), suggesting the PFC is involved in RCJs, but the type of retrieval task may moderate its role in RCJ performance. In addition to the PFC, regions of the parietal cortex have also been implicated in RCJ processes (Davidson et al., 2008; Simons et al., 2010). Neuropsychological studies show that patients with focal parietal lesions have reduced confidence in their memories as reflected by fewer ‘remember’ responses in remember/know paradigms (Davidson et al., 2008), and those with bilateral parietal lesions also have lower RCJ ratings in old/new recognition and source recollection tasks (Simons et al., 2010), suggesting the parietal cortex plays a role in the magnitude of RCJs.

Thus, it appears the PFC and parietal cortex play important roles in RCJs, and neuroimaging research has further identified subregions in which greater activity is correlated with low-confidence, high-confidence and RCJ accuracy. FMRI experiments have shown greater activation in the DLPFC and ventrolateral PFC for low RCJs as compared to high RCJs, using a face-name associative recognition paradigm (Chua et al., 2009a), item and source memory judgments (Hayes, Buchler, Stokes, Kragel, & Cabeza, 2011) and during an old/new verbal recognition task (Henson et al., 2000). Hayes et al. (2011) also showed that dorsal parietal regions showed greater activity during low confidence trials. This
increased activity with lower confidence may reflect greater demands on monitoring processes during low confidence trials (Chua, Pergolizzi, & Weintraub, 2014; Henson et al., 2000).

In contrast, some regions show greater activity with higher as compared to lower RCJ ratings. Several fMRI studies have shown that activity in the posterior parietal cortex (PPC) varies with the subjective level of the confidence judgment (Chua et al., 2006, 2009a; Chua, Schacter, & Sperling, 2009b; Kim & Cabeza, 2007). Greater activity in ventral parietal regions has also been correlated with high confidence responses, in contrast to dorsal regions which are more active with low confidence, suggesting subregions of the parietal cortex may differentially track magnitude of confidence ratings (Hayes et al., 2011). This is consistent with research demonstrating a distinction between the roles of superior and inferior parietal regions in other memory processes. Activity in the superior parietal lobe has been associated with top-down allocation of attention to memory retrieval (Ciaramelli, Grady, & Moscovitch, 2008; Olson & Berryhill, 2009), such as during the process of post-retrieval monitoring, and studies have shown activity in this region is greater when monitoring more uncertain or ambiguous targets: activity is greater for low vs. high confidence, for high-frequency vs. low frequency words (presumably because more effort is required to avoid false-alarms to more familiar words), and while evaluating lures that resemble studied material (Ciaramelli et al., 2008). Thus, it appears that the superior parietal cortex may be broadly involved in top-down direction of attention to the contents of memory retrieval (Ciaramelli et al., 2008; Sestieri, Shulman, & Corbetta, 2017), which is consistent with studies demonstrating its increased activation with low RCJs. In contrast, increased activity for high vs. low confidence in more ventral regions of the PPC is consistent with previous findings that the inferior PPC is more active during retrieval based on recollection than familiarity (Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Wagner, Shannon, Kahn, & Buckner, 2005), during retrieval success as compared to correct rejections (Konishi, Wheeler, Donaldson, & Buckner, 2000), and that PPC activity increases with the precision with which features of a visual stimulus are remembered (Richter, Cooper, Bays, & Simons, 2016), suggesting that increased inferior PPC activity with high confidence may reflect the strength of the memory being assessed (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Rugg & King, 2017).
In addition to correlational evidence implicating the parietal cortex in RCJs, non-invasive brain stimulation experiments have also provided evidence for a causal role of parietal regions in metamemory judgments: transcranial magnetic stimulation (TMS) to the inferior parietal lobule has been shown to selectively disrupt subjective but not objective measures of episodic memory retrieval (Sestieri, Capotosto, Tosoni, Luca Romani, & Corbetta, 2013).

Turning to RCJ accuracy, one fMRI study showed increased activity in the right aPFC with increased correlation between RCJ ratings and memory performance (Yokoyama et al., 2010). This finding is consistent with the aforementioned hypotheses regarding the possible hierarchical organization of metamemory functions within the PFC, wherein metamemory accuracy is reliant upon anterior PFC regions (Badre, 2008; Ryals et al., 2016).

There is some evidence, however, that both prefrontal and parietal activity associated with RCJs may vary based on the cognitive basis of the judgment (Kim & Cabeza, 2007). One fMRI study compared activity during high and low confidence ratings based on true and false memory, and showed greater activity in frontal (BA6) and lateral parietal regions (BA7) for low vs. high confidence during true recognition, but the opposite effect for false recognition (although it should be noted that for high vs. low confidence in false recognition, frontal activity was slightly more anterior [BA8] and lateral parietal activity was more anterior and ventral [BA39/40] compared to activity seen for low vs. high confidence in true recognition (Kim & Cabeza, 2007)). Because false recognition is thought to rely on familiarity, whereas true recognition is based on recollection of details, it may be that frontoparietal activity associated with RCJs varies with the information on which the judgment is based (Chua et al., 2014; Kim & Cabeza, 2007). Given the evidence that the cognitive basis of RCJs may be a combination of internal experience-based and external information-based cues (Koriat et al., 2008), the contribution of the parietal cortex to RCJs may also depend on whether individuals base confidence on internal vs. external information, or experience a conflict between these sources of information. Indeed, there is some evidence that the parietal cortex is involved in processing conflicts between internal and external cues in episodic memory tasks (O’Connor, Han, & Dobbins, 2010). One fMRI study found that when subjects were given cues from reliable and unreliable sources about the likelihood that a target was old or new, greater activity in the inferior parietal lobule was associated with invalid as compared to valid cueing, suggesting this region
may be recruited when external information-based cues conflict with experience-based cues during recognition (O’Connor et al., 2010). Experiment 2 uses fMRI to investigate the neural bases of RCJs based on valid and invalid cues about question difficulty in a semantic recognition task, to determine whether frontoparietal contributions to RCJs vary with the cognitive basis of the judgment.

Transcranial Direct Current Stimulation

Neuroimaging and lesion studies have provided some evidence about the brain regions involved in making accurate metamemory judgments (Kao et al., 2005; Ryals et al., 2016; Yokoyama et al., 2010). However, neuroimaging research is correlational and does not allow for interpretations about causal roles of brain regions in cognitive functions, and lesion studies are often limited by significant between-subject variations in lesion size and location (Pannu & Kasznia, 2005; Souchay et al., 2003). Transcranial direct current stimulation (tDCS), a form of non-invasive brain stimulation, has gained considerable attention in recent years as a potential intervention for a variety of clinical disorders, and a means by which to test causal roles of brain regions in cognitive functions (Colzato, Sellaro, & Nitsche, 2017; Flöel, 2014; Nitsche et al., 2008; Zhao et al., 2017), and has been shown to alter performance in various cognitive tasks, including metamemory performance (Chua & Ahmed, 2016; Chua, Ahmed, & Garcia, 2017). Given that neuroimaging research has provided us with regions of interest to use as targets for stimulation (Do Lam et al., 2012; Kao et al., 2005), tDCS may serve as a useful tool for investigating causal roles of PFC regions in JOL processes, and determining whether stimulation may serve as an effective way to enhance metamemory accuracy; this method will be used in Experiment 1.

Conventional tDCS involves passing a weak electrical current from one stimulating electrode, often referred to as the anode, to a return electrode, often referred to as the cathode, forming a circuit (Datta et al., 2009). A typical experimental protocol involves application of 1mA to 2mA of continuous direct current for 10 to 20 minutes (Flöel, 2014). Although electrical stimulation has been shown to alter neural excitability (Bikson et al., 2004; Hampstead, Brown, & Hartley, 2014; Stagg & Nitsche, 2011), some questions remain about the underlying biological mechanisms responsible for tDCS-induced changes (Stagg & Nitsche, 2011). There is evidence that the weak current that extends through the scalp and skull to reach the cortex may reduce the resting potential of neurons under the anode, and this
Depolarization results in an increase neuronal excitability (Bikson et al., 2004). Conversely, the current may raise the resting potential of neurons under the cathode, which decreases neuronal excitability (Bikson et al., 2004; Bikson, Rahman, & Datta, 2012). Research has suggested that tDCS may modify spontaneous neural activity by altering the conductance of ion channels to change polarization of the resting membrane potential (Stagg & Nitsche, 2011), as well as increasing astrocytic calcium levels that are thought to contribute to synaptic plasticity (Monai et al., 2016). There is also evidence that tDCS modulates neuronal firing after active current application has ceased, and that the mechanisms underlying these “after-effects” may differ from those associated with on-line stimulation (Stagg & Nitsche, 2011). Similarly to on-line tDCS effects, long-term effects of stimulation appear to depend on alterations in neuronal membrane potentials; however, tDCS has also been shown to induce synaptic modifications that lead to long-term synaptic plasticity after stimulation has ended (Stagg & Nitsche, 2011). Small increases in calcium concentration have been shown to induce long-term depression, while large rises in calcium lead to long-term potentiation, and these changes are thought to be dependent on activation of NMDA receptors resulting from stimulation (Stagg & Nitsche, 2011).

There is some skepticism about whether tDCS produces effects on the brain in practice (Horvath, Forte, & Carter, 2015a, 2015b), but several studies combining tDCS with neuroimaging techniques have provided evidence that stimulation does indeed produce changes in cortical activity, both at a local level and in regions functionally and anatomically connected to regions under the electrodes (Hampstead et al., 2014; Keeser et al., 2011). In one fMRI study, subjects receiving “anodal stimulation” over the parietal cortex, with the cathode over the PFC, showed greater activation in parietal regions compared to subjects receiving the opposite anodal/cathodal placement, as well as greater effective connectivity between parietal and frontal regions. Such results suggest that tDCS not only increases activity in the region under the stimulating electrode, but also modifies connectivity between the target region and a more distributed network (Hampstead et al., 2014). Another tDCS-fMRI experiment showed that “anodal stimulation” to the PFC altered resting-state network connectivity in regions both close to and distant from the site of stimulation, further suggesting tDCS directly modulates cortical activity, and this effect extends to areas connected to the target area (Keeser et al., 2011). One simultaneous tDCS-EEG study showed that compared to sham stimulation, “anodal” and “cathodal” tDCS over the sensorimotor cortex produced
changes in both spontaneous cortical activity, and event-related synchrony, providing evidence that tDCS directly alters rhythmic cortical activity as well (Roy, Baxter, & He, 2014). The timecourse of tDCS effects (online vs. offline) may also interact with the location of excitability (local vs. global): in one study, researchers applied TMS over the left posterior parietal cortex before, during, and after tDCS to the right posterior parietal cortex, and measured changes in excitability using EEG (Romero Lauro et al., 2014). By using EEG to record TMS-evoked potentials, a measure of cortical reactivity, the authors showed that tDCS induced a rise in global excitability both during and after tDCS, while local excitability was increased immediately following tDCS (Romero Lauro et al., 2014). Thus, although some questions remain about the exact mechanisms by which it exerts its effects (Stagg & Nitsche, 2011), tDCS has been shown to produce measurable changes in excitability at both the local and global level, both during and after application of stimulation (Hampstead et al., 2014; Keeser et al., 2011; Romero Lauro et al., 2014; Roy et al., 2014).

Evidence that tDCS alters brain activity in distributed regions outside the area under the electrodes has implications for the experimental design and interpretation of results from tDCS studies. For instance, it is possible that changes in cognitive processes can be achieved by stimulating multiple brain regions, or that stimulation to one region will have downstream effects on regions outside the sites of interest. Thus, it is important to recognize that tDCS has relatively low spatial focality, and studies aiming to determine the causal roles of specific cortical regions should acknowledge the possibility that areas outside the region of interest may contribute to changes in cognitive performance.

TDCS is relatively inexpensive and safe, with few side effects, making it a useful technique for determining causal roles of brain regions in numerous cognitive processes (Flöel, 2014; Jacobson, Koslowsky, & Lavidor, 2012). Most subjects experience a mild itching or burning sensation during stimulation, but this diminishes quickly (Woods et al., 2016). In order to control for placebo effects that may result from the experience of these scalp sensations, tDCS protocols typically employ a “sham” condition, in which current ramps up to the full amount of current delivered during active stimulation, and then immediately back down to baseline for the duration of the task (Gandiga, Hummel, & Cohen, 2006). Although tDCS blinding can be difficult to achieve (Davis, Gold, Pascual-Leone, & Bracewell, 2013; Horvath, Carter, & Forte, 2014), this condition is thought to minimize the potential effects of expectations
on behavioral results between stimulation groups, and one double-blind study showed no significant differences between sensations reported by participants in active and sham conditions (Gandiga et al., 2006).

However, using conventional tDCS, inferences about specific brain regions being targeted are limited by poor spatial resolution, resulting in possible stimulation of regions outside the area of interest, as evidenced by findings that the effects of stimulation extend beyond the target region (Hampstead et al., 2014; Keeser et al., 2011; Romero Lauro et al., 2014; Roy et al., 2014). In order to improve spatial resolution of tDCS, researchers have begun using “high definition” tDCS (HD-tDCS), in which smaller electrodes are positioned with one stimulating electrode among an array of four return electrodes. This montage allows for more focal stimulation of cortical regions of interest, and thus allows for stronger conclusions about specific causal structure-function relationships (Datta et al., 2009; Villamar et al., 2013). Although fewer studies have been conducted using HD-tDCS due to its relative novelty, there is some evidence it may produce stronger and longer-lasting changes in the brain as compared to conventional tDCS (Kuo et al., 2013), and reduce the likelihood of stimulating outside regions of interest. This increased specificity of stimulation, in addition to evidence of its ability to alter performance in other types of metamemory monitoring tasks (Chua & Ahmed, 2016; Chua et al., 2017), make HD-tDCS an ideal method to test how precise regions of the brain contribute to metamemory processes.

Given that accurate metamemory monitoring is crucial to effective learning and decision-making (Cosentino, Metcalfe, Cary, De Leon, & Karlawish, 2011; Metcalfe, 2002), and metamemory deficits have been identified in a wide range of clinical populations (Pannu & Kaszniak, 2005), it is important to understand the neurocognitive underpinnings of metamemory monitoring processes during both encoding and retrieval. Although metamemory judgments are influenced by information available at the time of the judgment, which may or may not be correlated with objective memory performance (Benjamin et al., 1998; Koriat, 1997; Koriat & Levy-Sadot, 1999), it remains unknown how the roles of brain regions associated with metamemory processes may differ with the cognitive basis of these judgments. By using neuroimaging and non-invasive brain stimulation methods, we aim to provide converging evidence to support the novel hypothesis that the way in which the brain supports metamemory monitoring during both encoding and retrieval varies with the cues on which judgments are based, which has previously
remained untested. Experiment 1 will use HD-tDCS to test the effects of stimulation on JOLs based on different sources of information to determine whether the causal roles of PFC subregions change with the cognitive bases of the judgments, and contribute to our understanding of the functional organization of the PFC as a whole. Experiment 2 will use fMRI to test how brain activity associated with RCJs during a semantic recognition task varies based on the presence and validity of external cues about question difficulty, to test how the brain supports recognition confidence judgments based on internal memory cues vs. external sources of information. Together, these experiments aim to elucidate how the neural mechanisms underlying metamemory monitoring during both encoding and retrieval may differ with the information on which subjective judgments are based, and inform structure-function relationships necessary to develop effective interventions to benefit those with metamemory deficits.
Chapter 2

Transcranial direct current stimulation to the prefrontal cortex alters encoding and judgments of learning based on fluency
Abstract

Past research has shown that judgments of learning (JOLs), subjective confidence judgments made at study about the ability to remember given items at test, are inferential in nature and based on cues available during encoding. Subjects tend to give higher JOLs to more fluently encoded items, despite having better recognition memory for less fluently encoded items, which leads to poor JOL accuracy. Research has implicated the dorsolateral and anterior regions of the prefrontal cortex (DLPFC and aPFC, respectively) in JOL and encoding processes, but no studies to date have tested how the roles of these regions vary with the cue on which JOLs are based. Here we used high definition transcranial direct current stimulation (HD-tDCS), a form of non-invasive brain stimulation, to test the causal roles of the DLPFC and aPFC in encoding success, JOL ratings, and JOL accuracy when subjects studied words that varied in their fluency at encoding. HD-tDCS over the DLPFC impaired encoding, as evidenced by an increase in subsequent false alarms. Turning to JOLs, we showed that for words that are less fluently encoded, the aPFC contributes to accurate JOLs by biasing individuals toward basing their subjective judgments on disfluency, which is predictive of objective memory success. Conversely, DLPFC and aPFC stimulation decreased JOL accuracy for high-frequency words, suggesting the roles of these regions in JOLs vary with the cognitive bases of the judgments. These results contribute to our understanding of the causal roles of subregions of the prefrontal cortex in objective and subjective memory processes, and how their contributions to metamemory accuracy vary with the cognitive basis of the judgments.
Introduction

A student studying for an exam must be able to monitor the status of her memory in order to guide her behavior: has she learned the material well enough to stop studying? If not, how much longer will she need to study in order to remember the material during the exam the next day? The ability to produce accurate judgments about the status of one's own memory, known as metamemory monitoring, is crucial to effective learning and decision-making (Metcalfe, 2002). One of the most commonly studied metamemory monitoring tasks, the Judgment of Learning (JOL) requires individuals to predict, during encoding, whether or not studied stimuli will be successfully remembered at test (Nelson & Dunlosky, 1991; Nelson & Narens, 1990). How well the predictive judgments correlate with actual memory performance is known as JOL accuracy. Research on the cognitive bases of JOLs has suggested these judgments are not based on direct access to a memory representation, but instead are inferential in nature and based on a variety of information, often referred to as cues, that are available at study (Koriat, 1997). Unfortunately, individuals often base their JOLs on cues that are not predictive of memory performance, resulting in inaccurate JOLs. Although research has shown that the accuracy of JOLs depends on the information on which the judgments are based, less is known about the brain regions that contribute to JOL processes. Regions of the prefrontal cortex (PFC), including the dorsolateral PFC (DLPFC) and anterior PFC (aPFC) have been implicated in the magnitude JOL ratings and their accuracy in predicting future retrieval (Do Lam et al., 2012; Kao et al., 2005; Ryals et al., 2016), but it remains to be understood how the roles of these regions may vary with the cognitive basis of the judgment. The current study aims to test how the PFC supports JOL accuracy by testing whether the roles of the DLPFC and aPFC may vary when JOLs are based on information that is and is not predictive of actual memory performance.

A significant body of research has been conducted on the cognitive bases of JOLs, and it is generally accepted that people make inferences about the memorability of given stimuli based on various cues (Benjamin et al., 1998; Koriat, 1997). Typically, individuals tend to use “intrinsic cues,” which are characteristics of stimuli, for example factors related to encoding fluency such as item difficulty or speed of processing, to make their JOLs, despite the fact that these factors tend not to influence memory performance (Begg et al., 1989; Koriat, 1997; Koriat & Ma’ayan, 2005). For example, in a number of
studies, ease of processing, as measured by encoding speed, positively correlated with JOLs, but not subsequent memory performance (Begg et al., 1989; Benjamin et al., 1998; Hertzog et al., 2003). Similarly, studies manipulating word frequency, which also varies ease of processing, showed that subjects gave higher JOLs to high-frequency words, but actually have better memory for low-frequency words as evidenced by higher hit rates and fewer false alarms (Benjamin, 2003). Thus, individuals often make incorrect inferences about the future memorability of items because they rely too heavily on intrinsic cues that do not always predict memory performance.

Not only have individuals been shown to rely on misleading cues when making JOLs, but research has also shown they tend to undervalue cues that are predictive of future memory performance (Shaughnessy, 1981; Sungkhasettee et al., 2011). These “extrinsic cues” are related to the conditions of learning, such as stimulus repetition, and duration and depth of encoding (Dunlosky & Matvey, 2001). For example, participants studied paired-word associates using either interactive imagery or rote rehearsal strategies, and made an immediate JOL following each pair (Shaughnessy, 1981). There was no difference in JOLs for the imagery versus rote rehearsal strategies, but memory performance was greater for pairs encoded using imagery (Shaughnessy, 1981). Similarly, another study showed that participants’ JOLs did not differ for words that were studied upright or inverted 180 degrees, but participants later recalled more inverted than upright words (Sungkhasettee et al., 2011). This effect persisted even when participants were given several study-test cycles, which provided them with the opportunity to potentially learn about the benefits of deeper processing of inverted words. Such results suggest subjects discount the memory benefit of ‘desirable difficulties’, such as the increased processing required to encode inverted words, which results in better subsequent memory performance (Sungkhasettee et al., 2011).

In addition to understanding what kinds of cues participants do and do not use to make JOLs, it is also critical to determine how JOLs relate to subsequent memory performance, which is referred to as JOL accuracy (Benjamin & Diaz, 2008). Unfortunately, immediate JOLs, those made during or immediately after study, are typically not very accurate (Nelson & Dunlosky, 1991) because participants often based JOLs on fluency of encoding, which is nondiagnostic of memory success, or fail to consider the depth of encoding when making JOLs (Begg et al., 1989; Benjamin, 2003; Koriat, 1997;
Shaughnessy, 1981; Sungkhasettee et al., 2011). The current experiment manipulates the cognitive basis of JOL accuracy by testing how JOLs vary for words of different fluencies based on familiarity with the stimulus (i.e., high vs. low frequency words) and fluency of perceptual processing (i.e., words presented in an upright or inverted orientation), and investigates how the brain supports JOL accuracy when judgments are made on cues that vary in their influence on memory performance.

Turning to the brain bases of JOLs, neuroimaging and lesion studies provide converging evidence that the prefrontal cortex (PFC) is important for JOLs, with different subregions relating to JOL ratings and JOL accuracy (Kao et al., 2005; Vilkki et al., 1999). Lesion studies have focused more broadly on the prefrontal cortex and showed that, compared to posterior lesion and control patients, patients with prefrontal lesions made less accurate predictions at the time of study about later retrieval success (Vilkki et al., 1999). It is worth noting that there may be some hemispheric differences, with damage to the left frontal cortex leading to over-predicting recall ability, and damage to the right frontal cortex leading to both under-prediction and over-prediction of recall ability (Vilkki, Servo, & Surma-aho, 1998).

Neuroimaging experiments have identified subregions of the PFC that may be responsible for specific aspects of JOLs, with some between study variations due to stimulus type and paradigm (Do Lam et al., 2012; Kao et al., 2005). For JOL ratings, greater activity in the medial PFC and orbitofrontal cortex [Brodmann areas (BA) 10 and 11] was correlated with higher JOL ratings in a face-name associative encoding task (Do Lam et al., 2012), whereas there was greater activity in lateral (BA 44/6) and ventromedial prefrontal (vmPFC; BA11) regions for “will remember” predictions as compared to “will forget” predictions a scene encoding task (Kao et al., 2005), regardless of JOL accuracy. Similarly, in a verbal encoding task, Yang et al. (2015) found that “will remember” predictions were correlated with greater activity in the DLPFC (BA8) and vmPFC (BA10), as compared to “will forget” predictions. ERP studies have also shown that the process of making JOLs produces a medial frontal positive waveform, the magnitude of which is correlated with the magnitude of JOLs ratings, implicating the medial PFC in the cognitive process of making JOLs as well as the subjective level of confidence expressed (Müller et al., 2016; Skavhaug et al., 2010; Skavhaug, Wilding, & Donaldson, 2013). Taken together, these findings suggest that both anterior (e.g., orbitofrontal), and posterior (e.g. DLPFC) regions within the PFC may track the magnitude of the JOL ratings, and anterior regions associated with JOLs are more medial, while
posterior regions are more lateral. In terms of JOL accuracy, individual differences in JOL accuracy, as indexed by Gamma coefficients, were correlated with vmPFC (BA11) activity (Kao et al., 2005). Although there is some variation in the subregions whose activity correlates with JOLs, both anterior and posterior prefrontal subregions have been shown to play roles in JOL ratings or accuracy, and are regions of interest that will be examined in the current experiment.

Indeed, one brain stimulation study compared more anterior and more posterior prefrontal regions in JOLs (Ryals et al., 2016). In an associative recognition and JOL task, continuous theta burst stimulation to the aPFC improved JOL accuracy as compared to DLPFC and vertex stimulation, suggesting a causal role of the aPFC in making accurate judgments (Ryals et al., 2016). Although the main effect of stimulation location on trial-by-trial measures of JOL accuracy was marginal, aPFC stimulation decreased overall JOL ratings for subsequent misses, suggesting stimulation made subjects more aware of failures during encoding that would result in poor memory performance at test. This raises the hypothesis that the aPFC supports JOL accuracy by increasing sensitivity to information at encoding that is predictive of later memory performance (Ryals et al., 2016).

The results of Ryals et al.'s (2016) experiment testing the casual roles of the aPFC and DLPFC in JOLs showed no significant effect of stimulation on recognition performance. This is notable because past research has implicated the PFC in successful episodic encoding. Although there is some evidence that patients with DLPFC lesions tend to be more impaired in recall than recognition (Mangels, Gershberg, Shimamura, & Knight, 1996; Staresina & Davachi, 2006), other studies have shown that patients with DLPFC lesions show impaired episodic recognition (Duarte, Ranganath, & Knight, 2005; Wheeler, Stuss, & Tulving, 1995), and neuroimaging research has more specifically highlighted the role of the left DLPFC in successful episodic encoding as measured by subsequent recognition (Blumenfeld & Ranganath, 2006, 2007; Sperling et al., 2001). Given that common brain regions, such as the DLPFC, are associated with both objective and subjective memory processes, it is important to distinguish between the neural underpinnings of JOLs as compared to successful encoding. Some metamemory studies have shown a dissociation between regions that contribute to JOLs and encoding: patients with prefrontal lesions show impaired JOLs but intact memory (Vilkki et al., 1999), and one fMRI study showed that activity in the medial PFC and aPFC was associated with the process of making JOLs, even when
exclusively masking JOL-related activity with contrasts reflecting encoding-related activity (i.e. identifying JOL-related activity that is independent of encoding) (Do Lam et al., 2012). ERP research has also shown that JOLs and encoding processes produce similar positive waveforms over fronto-central and parieto-occipital sites in early time windows, but that these signals diverge at later time windows, with a negative waveform associated with JOLs but not encoding over left central sites (Skavhaug et al., 2010). Taken together, these results suggest that the neural mechanisms underlying JOLs and episodic encoding are partially dissociable. Understanding the shared and distinct roles of brain regions in subjective and objective memory is crucial to clarifying the neural bases of JOL accuracy, which reflects the relationship between these two processes. Given that the aim of the current study is to test the role of the PFC in JOLs based on cues that vary in their diagnosticity of memory performance, a major focus is how the aPFC and DLPFC may differentially contribute to encoding and JOL processes.

The current experiment uses non-invasive brain stimulation to test the novel hypothesis that the aPFC may bias individuals to make JOLs based on information that is predictive of later memory success, and discount cues that are nondiagnostic of future performance (Ryals et al., 2016). If the aPFC increases sensitivity to diagnostic over nondiagnostic cues, aPFC stimulation should result in increased JOL accuracy across cue types, as it will bias individuals toward basing JOLs on information that is predictive of memory success. Gaining knowledge about the causal role of the aPFC in JOL accuracy, and whether its role varies with the cognitive basis of the judgment, is a crucial step toward understanding the neural underpinnings of accurate memory awareness, establishing precise structure-function relationships, and informing future interventions to treat metamemory impairments.

To test the roles of the aPFC and DLPFC in JOLs, we used transcranial direct current stimulation (tDCS), a non-invasive brain stimulation technique, to manipulate brain activity. Conventional tDCS involves passing a weak electrical current from one stimulating electrode, typically referred to as the anode, to one return electrode, typically referred to as the cathode. Because conventional tDCS uses relatively a relatively large anode and cathode that are spaced farther apart, inferences about specific brain regions being targeted are limited by poor spatial focality, resulting in possible stimulation of regions outside the area of interest (Datta, Truong, Minhas, Parra, & Bikson, 2012; Hampstead et al., 2014; Keeser et al., 2011). To improve the spatial focality of tDCS, the current experiment uses “high definition”
tDCS (HD-tDCS), in which smaller electrodes are positioned with one stimulating electrode among an array of four return electrodes. Although fewer studies have been conducted using HD-tDCS due to its relative novelty, there is evidence it produces changes in more focal regions of the cortex (Kuo et al., 2013), with stimulation constricted to the area within the radius of the cathodes (Villamar et al., 2013). This increased spatial specificity reduces the likelihood of stimulating outside targeted regions of interest, and makes HD-tDCS a better method for testing how more precise subregions of the PFC contribute to JOL accuracy. Furthermore, HD-tDCS has been useful at identifying the causal role of the DLPFC in a different metamemory task, namely the feeling-of-knowing task, using a semantic retrieval paradigm (Chua & Ahmed, 2016; Chua et al., 2017). The current experiment uses HD-tDCS to test the causal roles of the aPFC and DLPFC in JOL processes, and whether their roles vary with the cognitive basis of the judgments.

Methods

Participants

Twenty-five healthy Brooklyn College students were consented to participate in this six-session study for financial compensation ($15/hour for 4.5 hours). One participant only completed one session and was withdrawn from the study due to poor impedance at the second session. Thus, data from 24 participants (12 F, ages 18-26, M = 21.0, SD = 2.11 years) were included. G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) was used to determine that for a repeated measures ANOVA with 1 group and 3 measurements, a sample size of 24 subjects was needed for 80% power and a moderate effect size of 0.27, which is lower than the effect size reported in a similar tDCS experiment testing the roles of DLPFC and anterior temporal lobe (ATL) in metamemory (Chua & Ahmed, 2016). All participants were right-handed, learned English before the age of 5, and were free from any self-reported neurological or psychological disorders, medical or skin conditions, unhealed wounds on the scalp, neck, face or forehead, and metallic implants. All participants gave written consent in a manner approved by the Human Research Protection Program at the City University of New York.
HD-tDCS Protocol

High definition transcranial direct current stimulation (HD-tDCS) was delivered using the Soterix 1x1 tDCS device (Model 1224-B, Soterix Medical, New York, NY) connected to the Soterix 4x1 adapter (Model 4X1-C3 and Model 4X1-C3A, Soterix Medical, New York, NY). All stimulation conditions used five sintered Ag/AgCl ring electrodes (12 mm outer radius and 6 mm inner radius) to deliver low current stimulation. In a within-subjects design, each participant received 3 stimulation sessions: active stimulation over the left dorsolateral prefrontal cortex (DLPFC) and the anterior prefrontal cortex (aPFC), and sham stimulation over the left parietal cortex. Sham stimulation was administered over a different site from active stimulation to minimize the likelihood that subjects would be able to distinguish sham from active stimulation (i.e., subjects knew one of the three sessions would involve sham stimulation, and applying two sessions to the same site would indicate that one was sham, as well as possibly allow subjects to directly compare sensations experienced during both sessions at the same location).

Electrode configurations were determined using Soterix HD-Explore (Soterix Medical, New York, NY), with electrode locations corresponding to locations on the 10-20 EEG system. For aPFC stimulation, the anode was placed over FPZ with return electrodes at FP1, FP2, AF3, and AF4 (Fig 2A). For left DLPFC stimulation, the anode was placed over F3 with return electrodes at AF3, F1, F5 and FC3 (Fig. 2B). For sham, the anode was placed over CP3 with return electrodes at C3, CP5, CP1, and P3 (not shown).

Before each stimulation session, subjects received a “pre-stimulation tickle”, which involved a 30-second ramp up to 1 mA and immediate 30-second ramp down to baseline, to familiarize participants with the sensation of stimulation at half the current intensity of full stimulation, and determine whether they were able to tolerate stimulation. All participants were able to tolerate stimulation.

During each active stimulation session, following a 30-second ramp up, 2 mA of current was delivered through the stimulating electrode, often referred to as the anode, and distributed equally among the four return electrodes (0.5 mA each), often referred to as cathodes, for the duration of the study task. During sham stimulation sessions, current ramped up to 2 mA and then down to 0.1 mA over the course of 30 seconds; current was maintained at 0.1 mA for the duration of the task, which is an insufficient strength of current to produce any cognitive effects (Gandiga et al., 2006). The ramp up and down at the start of the sham session mimics the skin sensations experienced by subjects during active stimulation.
(e.g., itching), and has been shown to serve as an effective control in conventional tDCS so participants have difficulty distinguish between active and sham stimulation (Gandiga et al., 2006, but see Davis, Gold, Pascual-Leone, & Bracewell, 2013; Horvath, Carter, & Forte, 2014; O’Connell et al., 2012). Stimulation began ~3 min before start of the encoding task, and was aborted upon completion of the task.

Each participant completed 6 sessions, which consisted of 3 study/test cycles, with 1 week between each. Study and test were separated by 24 hours to ensure that there were no residual HD-tDCS effects at test. Sessions were counterbalanced for stimulation sites and word lists across participants.

Stimuli and Procedure

Over 3 study/test sessions, participants studied a total of 600 words (average length: 5.63 letters, SD = 1.5) from the MRC Psycholinguistics Database (Wilson, 1988), which were divided into 3 study sets (100 words per set) and 3 test sets (200 words per set; 100 old/100 new). 50% of the words were presented upright, 50% presented inverted 180 degrees, and of these, 50% were high-frequency and 50% low-frequency words. Word frequency, which refers to the number of times a particular word appears in a given corpus, was determined using the Kucera-Francis Frequency Scale (Kučera, H., & Francis, W. N., 1967). Low frequency words were defined as those with values between 1 and 5, and high-frequency words ranged from 50-492, based on past research using Kucera-Francis Frequency ratings to define high vs. low-frequency words (Diana & Reder, 2006; Rudell, 1993). Words were matched across conditions (frequency and orientation) and between lists by word length, concreteness, and imageability.

Study Task: The study task began ~3 minutes after stimulation began. During the first visit, this time was filled with a practice session, during which participants studied and gave JOLs to 6 words to familiarize them with the task. During visits 2 and 3, participants were at rest during the 3 minutes between the start of stimulation and start of the study task. All stimuli were presented using Psychopy v1.84.0 (Peirce, 2007).

During each study session, subjects were presented with 100 total words for 2.5 seconds each. There were 25 words per frequency/orientation condition: upright/high frequency, inverted/high frequency,
upright/low frequency, and inverted/low frequency. Immediately after presentation of each word, participants saw a rating scale from 0-100% in 10% increments and gave a JOL rating (2.5 seconds) to indicate the likelihood they would remember the previously displayed word 24 hours later at test. Participants gave JOLs using the number scale on the keyboard from left to right, such that “~” = 0% confidence, 1 = 10% confidence, 2 = 20%, 3 = 30%... with 0 (i.e., where ‘10’ would be) indicating 100% confidence. Study task duration was 8.3 min, excluding two optional 1 min breaks and self-paced instructions.

After each study session, participants were given a post-stimulation questionnaire consisting of possible side effects they may have experienced (i.e. headache, neck pain, scalp pain, tingling, burning sensation, skin redness, sleepiness, trouble concentrating, acute mood changes, other), based on guidelines for reporting tDCS effects proposed in previous literature (Brunoni et al., 2011). Subjects indicated whether they experienced any of the listed side effects on a scale from 1-4 (1-absent, 2-mild, 3-moderate, 4-severe), and whether they believed there was a relationship between the side effect and the stimulation on a scale from 1-5 (1-none, 2-remote, 3-possible, 4-probably, 5-definite). Finally, they were asked to indicate whether they thought they received active or sham stimulation (Appendix A).

Recognition Task: Twenty-four hours after each study session, subjects completed a self-paced old/new recognition test. Each test consisted of all 100 studied words, in the same orientation as presented at study, and 100 new words matched on frequency and orientation (25 per frequency/orientation condition). Subjects indicated their response via keypress (1 = old; 2 = new). Participants were compensated and debriefed after the final test on their last session. Debriefing included a brief explanation of the general hypotheses being addressed by the experiment, and disclosing which of the three stimulation sessions were active vs. sham.

Data Analyses

To assess memory, we examined the effects of cue type and stimulation on hits and false alarms separately. Past research has shown that in addition to depth of encoding, greater memory for high-frequency than low-frequency words may also be partly driven by false alarms to high-frequency words at test (Benjamin, 2003), so it was important to assess false alarms separately rather than in a composite
measure of memory such as corrected recognition or d’. Furthermore, we predicted that memory for inverted words would be better than for upright words due to depth of encoding, but perceptual fluency may also cause individuals to produce more false-alarms to upright words at test (Johnston, Dark, & Jacoby, 1985).

JOL accuracy was assessed using \( d_a \), a signal-detection based trial-by-trial measure of metacognitive accuracy, in which a metacognitive hit, such as giving a high confidence rating to a successfully remembered item, is plotted against the probability of a metacognitive false-alarm, such as a high-confidence item that is not successfully remembered. \( d_a \) is calculated by the following formula: 
\[
\sqrt[2]{\frac{y_0}{\sqrt{1+m^2}}}
\]
in which \( y_0 \) represents the y-intercept and \( m \) represents the slope of the isosensitivity function (Benjamin & Diaz, 2008; Masson & Rotello, 2009).

The relationships between recognition performance, JOL ratings, JOL accuracy, HD-tDCS site, word frequency, and word orientation were analyzed using mixed linear models and post-hoc t-tests in SPSS. Full factorial designs examining effects of frequency and orientation were not evaluated due to low trial counts for combined frequency, orientation, and performance conditions (e.g., some individuals had no high-frequency upright misses). Instead, separate models were used to analyze the effects of: 1) frequency and stimulation on recognition performance, JOL ratings, and \( d_a \), and 2) the effects of orientation and stimulation on recognition performance, JOL ratings, and \( d_a \).

Previous research has suggested the effects of tDCS may vary based on a number of individual differences, including anatomical differences related to sex, age, and head size (Datta et al., 2012). Furthermore, research has shown that internal psychological states, such as transient changes in mood (Harrison et al., 2008; Mayberg et al., 1999), alertness (Braboszcz & Delorme, 2011), and motivation (Berryhill, Peterson, Jones, & Stephens, 2014) alter baseline neural activity, which is likely to mediate the effects of tDCS (Berryhill et al., 2014; Learmonth, Thut, Benwell, & Harvey, 2015; Neuling, Rach, & Herrmann, 2013). Thus, in order to control for the effects of individual variation in responses to tDCS, we included variables reflecting participants’ sex, age, head size, education level, mood, alertness, and sensitivity to stimulation as covariates in these mixed linear models. To take into account the number of fixed effects parameters being estimated, all models used a restricted maximum likelihood procedure (SPSS version 23.0) to yield unbiased parameter estimates. All results were considered significant at
In order to determine whether the inclusion of individual differences improved the strength of our mixed model, we first constructed a model using stimulation site, word frequency, and subsequent performance as predictors of JOL ratings. Akaike’s information criterion (AIC) showed that relative to the model excluding covariates (AIC = 4734.677), the addition of covariates reflecting individual differences in stimulation experience (headache, scalp pain, tingling, burning, skin redness), mood state (sleepiness, trouble concentrating, acute mood changes), sex, age, and head circumference improved the strength of the model (AIC = 4646.953). A model using stimulation site, word orientation, and subsequent performance as predictors of JOL ratings also showed that relative to the model excluding covariates (AIC = 4752.218), the inclusion of covariates improved the strength the model (AIC = 4666.139). Models testing effects of stimulation and frequency, and stimulation and orientation on all other dependent variables (hits, false alarms, and $d_a$) showed that the strength of each model was improved by the inclusion of covariates. Therefore, all subsequent models included these measures of individual differences as covariates, and subject ID as a random effect. Stimulation site was coded with aPFC as the reference, and all other predictors were mean-centered to allow for interpretation of the intercept and avoid multicollinearity when assessing interactions. All covariates in each model were evaluated at values of 0. Outliers were determined with SPSS, using a step of 1.5 x Inter-quartile range, with quartiles defined according to Tukey’s hinges. Outlier analyses were conducted on overall JOL ratings, corrected recognition, Gamma coefficients, and $d_a$ performance. One subject was excluded from all analyses of JOL accuracy based on outlier $d_a$ performance during sham stimulation ($d_a = 0.99$), and another subject was excluded from analyses of JOL accuracy which included frequency as a predictor, due to extreme outlier performance for low-frequency words ($d_a = -3.29$).

**Results**

**Task Duration**

Because participants had optional breaks to take during encoding, and stimulation was terminated at the end of the encoding, we first evaluated whether there were differences in stimulation duration at encoding using a repeated measures ANOVA. Total study task duration ranged from 9.24 -
11.28 min (M = 9.62 min, SD = .55 min), and stimulation site did not significantly affect time to complete study \[F(2,46) = 0.340, p=0.713\]. We also tested whether or not the time to complete the recognition test (range: 4.61 – 10.88 min; M = 7.56 min; SD = 1.41) varied by stimulation condition, and there were no differences between stimulation conditions \[F(2,46) = 0.183, p=0.835\].

**Participant Blinding**

To assess whether participants were blinded to the stimulation condition (i.e., active vs. sham), we examined their responses on the post-stimulation sensation questionnaire about whether they thought they received active or sham stimulation. In the sham condition, 14 participants correctly guessed they received sham stimulation, and 10 incorrectly guessed they received active stimulation. In the DLPFC condition, 17 participants correctly guessed they received active stimulation, and 7 incorrectly guessed they received sham stimulation. In the aPFC condition, 17 participants correctly guessed they received active stimulation, and 7 incorrectly guessed they received sham stimulation. To test if the belief about stimulation differed between conditions, we used a repeated measures logistic regression with aPFC stimulation as a reference. Significantly more participants guessed they received active stimulation during aPFC stimulation as compared to sham (Wald Chi-Square = 5.549, p<0.05) and during DLPFC compared to sham (Wald Chi-Square = 3.953, p<0.05) but there was no significant difference in subjects’ guesses that they received active stimulation between DLPFC and aPFC sessions (Wald Chi-Square = 0.400, p=0.527).

**Memory Performance**

*Subsequent Recognition - Word Frequency*

In order to understand any potential effects of stimulation on recognition, we examined changes in hits and false alarms separately, and constructed two models with stimulation and frequency as predictors for each trial type (Fig. 3), and included individual differences as covariates. For hits (Fig. 3A), word frequency was a significant predictor \[F(1, 251.211) = 18.536, p<0.001\], with higher hit rates for low-frequency words (M = 0.660, SE = 0.038) compared to high-frequency words (M = 0.618, SE = 0.038; Mean Difference = 0.042, 95% CI [0.023, 0.061]), as expected due to deeper encoding of low-frequency
words. There was no main effect of stimulation site (p>0.370), and no significant interaction between stimulation site and word frequency (p>0.310).

Because the recognition benefit of low-frequency words has also shown to be driven by higher false alarms for high-frequency words (Benjamin, 2003), we analyzed the effects of stimulation and word frequency on false alarm rates (Fig. 3B), again including individual differences as covariates. There was a significant main effect of stimulation [F(2,254.108) = 5.270, p<0.01], a significant main effect of frequency [F(1,251.381) = 57.946, p<0.001], and no significant interaction between stimulation and frequency on mean predicted false alarms (p>0.79). Pairwise comparisons testing the significant main effect of stimulation showed that as compared to false alarm rates following sham stimulation (M = 0.241, SE = 0.027), DLPFC stimulation significantly increased subsequent false alarms (M = 0.267, SE = 0.027; Mean Difference = 0.026, 95% CI [0.002, 0.049], p<0.05). False alarms with DLPFC stimulation were also significantly higher than with aPFC stimulation (M = 0.230, SE = 0.027; Mean Difference = 0.037, 95% CI [0.014, 0.060], p<0.01), but there was no difference between aPFC and sham (p>0.38). There was also a significant main effect of word frequency: as expected, subjects had higher false alarm rates for high-frequency words (M = 0.279, SE = 0.027) than low-frequency words (M = 0.213, SE = 0.027; Mean Difference = 0.067, 95% CI [0.049, 0.084], p<0.01). Taken together, analyses of the effects of stimulation and word frequency on hits and false alarms suggest that subjects benefit from the distinctiveness of low-frequency words at encoding, as shown by higher hit rates for low- vs. high-frequency words, but encoding was impaired by DLPFC stimulation, as shown by an increase in subsequent false alarm rates.

**Subsequent Recognition - Word Orientation**

We next constructed a model to examine the effects stimulation and word orientation as predictors of hits and false alarms. A model including stimulation and orientation as predictors of hit rates with individual differences as covariates showed no significant main effect of stimulation [F(2,252.779) = 1.120, p>0.32], and no interaction between stimulation and orientation [F(2,251.186) = 0.528, p>0.59], but a significant main effect of orientation on predicted hit rates [F(1,251.186) = 43.962, p<0.001]. Hit rates for inverted words (M = 0.669, SE = 0.038) were significantly higher than hit rates for upright words.
(M = 0.609, SE = 0.038; Mean Difference = 0.061, 95% CI [0.043, 0.079]) (Fig. 4A), suggesting subsequent memory benefitted from greater depth of encoding and distinctiveness of inverted words at encoding.

Turning to false alarm rates, there was a significant main effect of stimulation site [F(2,254.243) = 5.025, p<0.01], a significant main effect of orientation [F(1,251.399) = 5.534, p<0.05], and no interaction between stimulation and orientation on predicted false alarm rates [F(2,251.399) = 0.024, p>0.97]. False alarms for upright words (M = 0.257, SE = 0.027) were significantly higher than false alarms for inverted words (M = 0.236, SE = 0.027; Mean Difference = 0.021, 95% CI [0.003; 0.039]) (Fig. 4B). Pairwise comparisons on the effects of stimulation on false alarms showed that relative to sham (M = 0.241, SE = 0.027) false alarms were significantly higher following DLPFC stimulation (M= 0.267, SE = 0.027; Mean Difference = 0.026, 95% CI [0.001, 0.050], p<0.05). False alarms following DLPFC stimulation were also significantly higher than aPFC stimulation (M = 0.230, SE = 0.027, Mean Difference = 0.037, 95% CI [0.013, 0.061], p<0.01), with no significant difference between sham and aPFC (p>0.39). Thus, similarly to the effects of stimulation on memory for high- and low-frequency words, DLPFC stimulation increased false alarm rates. Taken together, results suggest that encoding was poorer under DLPFC stimulation, as evidenced by a poorer ability to distinguish “new” and “old” items, as seen by an increase in overall false alarm rates after DLPFC stimulation regardless of frequency/orientation.

**Metamemory Performance**

*Mean JOL Ratings – Word Frequency*

A primary aim of the current experiment was to test the effects of prefrontal stimulation on the magnitude and accuracy of JOLs based on different cue types. Across all sessions, subjects’ mean JOLs ranged from 38.30 to 88.59 (M = 63.99, SD = 13.33). We first constructed a model using stimulation site, word frequency, and subsequent performance on mean JOL ratings with individual differences as covariates. There was a significant main effect of subsequent recognition [F(1,244.781) = 31.744, p<0.001] and a significant main effect of frequency [F(1,244.781) = 34.218; p<0.001]. There was no significant main effect of stimulation [F(2,247.462) = 0.297, p>0.74], and no 2- or 3-way interactions.
Subsequent recognition performance was a significant predictor of mean JOL ratings: as expected, mean JOLs for items subsequently remembered (M = 65.525, SE = 3.068) were significantly higher than mean JOLs for subsequent misses (M = 59.947, SE = 3.068; Mean Difference = 5.578, 95% CI [3.628, 7.528]). Word frequency was also a significant predictor of mean JOLs as predicted; JOLs for high frequency words (M = 65.632, SE = 3.068) were significantly higher than those for low frequency words (M = 59.841, SE = 3.068; Mean Difference = 5.791, 95% CI [3.841, 7.742]). Results are consistent with past work (Benjamin, 2003; Jia et al., 2016; Sungkhasetee et al., 2011), and support our hypothesis that subjects base JOLs on fluency at encoding, as reflected by higher JOL ratings for high-frequency words. Stimulation site did not predict differences in mean JOL ratings and there were no significant interactions in the effects of stimulation, frequency, and performance on mean JOLs.

Mean JOL Ratings – Word Orientation

We then constructed a model using stimulation site, word orientation, and subsequent performance on mean JOL ratings with individual differences as covariates. Again, as predicted, subjects gave higher JOLs to subsequent hits (M = 65.493, SE = 3.116) than subsequent misses (M = 60.350, SE = 3.116; Mean Difference = 5.143, 95% CI [3.316, 6.970], F(1,244.816) = 30.748, p<0.001). Consistent with research showing subjects often incorrectly base confidence on fluency of encoding (Johnston et al., 1985; Koriat, 1997), subjects also gave higher JOLs to upright words (M = 64.026, SE = 3.116) than inverted words (M = 61.817, SE = 3.116; Mean Difference = 2.210, 95% CI [0.383, 4.037], F(1,244.816) = 5.675, p<0.05). However, stimulation did not significantly predict mean JOLs [F(2,247.121) = 0.414, p>0.66] and there were no significant 2- or 3-way interactions. Taken together, results demonstrate typical findings regarding the cues used to make JOL ratings: subjects gave higher JOLs to high-frequency and upright words, suggesting they base confidence in retrieval on ease of encoding (Koriat, 1997), but stimulation had no effect on JOLs for any cue type.

JOL Accuracy – Word Frequency

As is typical with immediate JOLs, subjects’ JOL accuracy was poor to moderate, with mean $d_a$ values across all cue conditions ranging from -0.28 to 0.73 (M = 0.22, SD = 0.23), excluding outliers. To
understand how stimulation affected trial-by-trial measures of JOL accuracy, and whether this varied by word frequency, we constructed a mixed linear model using stimulation site and frequency as predictors of $d_a$, a signal detection based measure of metamemory accuracy (Benjamin & Diaz, 2008), and included individual differences as covariates. There was a main effect of word frequency on predicted mean $d_a$ 
$[F(1,490.084) = 14.756, p<0.001]$, with JOL accuracy being significantly higher for low-frequency words ($M = 0.258, SE = 0.055$) than high-frequency words ($M = 0.175, SE = 0.055$; Mean Difference = 0.083, 95% CI [0.040, 0.125]). There was also a main effect of stimulation site $[F(2,498.879) = 7.642, p<0.002]$, with worse JOL accuracy in the DLPFC condition ($M = 0.155, SE = 0.056$) compared to sham ($M = 0.229, SE = 0.057$; Mean Difference = -0.074, 95% CI [-0.033, -0.016], $p<0.05$) and compared to aPFC stimulation ($M = 0.265, SE = 0.056$; Mean Difference = -0.110, 95% CI [-0.168, -0.053], $p<0.001$), with no difference in $d_a$ between aPFC and sham conditions ($p>0.25$). This main effect of stimulation was qualified by a significant interaction between stimulation site and word frequency $[F(2,490.084) = 11.780, p<0.001]$, showing differential effects of stimulation depending on word frequency. For low frequency words, aPFC stimulation significantly improved $d_a$ values relative to sham (Mean Difference = 0.164, 95% CI [0.083, 0.245], $p < 0.001$), and relative to DLPFC stimulation (Mean Difference = 0.181, 95% CI [0.104, 0.259], $p<0.001$). In contrast, for high-frequency words, aPFC and DLPFC stimulation significantly impaired accuracy relative to sham (aPFC vs. sham: Mean Difference = -0.092, 95% CI [-0.173, -0.011], $p<0.05$); DLPFC vs. sham: Mean Difference = -0.132, 95% CI [-0.210, -0.054], $p<0.002$) (Fig. 5A). Taken together, these results suggest aPFC stimulation selectively improved JOL accuracy for the less fluent condition, while both prefrontal stimulation sites decreased accuracy for words that were more fluent at encoding.

**JOL Accuracy – Word Orientation**

To test whether stimulation affected JOL accuracy when judgments were based on word orientation, we constructed a mixed linear model using stimulation site and word orientation as predictors of $d_a$, including individual differences as covariates. There was a significant main effect of word orientation $[F(1,513.516) = 13.418, p<0.001]$, a significant main effect of stimulation site $[F(2,531.256) = .796, p<0.05]$, and a significant interaction between orientation and stimulation site $[F(2, 513.516) = 3.666,$
p<0.05] on predicted mean $d_a$. Predicted mean $d_a$ was higher for inverted words ($M = 0.260, SE = 0.041$) relative to upright words ($M = 0.156, SE = 0.041; \text{Mean Difference} = 0.104, 95\% \text{ CI [0.048, 0.160]}$). There was a significant main effect of stimulation site on JOL accuracy, and pairwise comparisons demonstrated that $d_a$ for aPFC stimulation ($M = 0.266, SE = 0.044$) was marginally higher than during sham stimulation ($M = 0.195, SE = 0.045; \text{Mean Difference} = 0.071, 95\% \text{ CI [0.010, 0.152]}, p<0.09$), and significantly higher than during DLPFC stimulation ($M = 0.162, SE = 0.044; \text{Mean Difference} = 0.104, 95\% \text{ CI [0.029, 0.179]}, p<0.01$). This main effect of stimulation was qualified by a significant interaction between stimulation site and word orientation on mean predicted $d_a$. Pairwise comparisons revealed that stimulation site had no effect on $d_a$ values for upright words. For inverted words, relative to sham ($M = 0.255, SE = 0.051$), aPFC stimulation significantly improved JOL accuracy ($M = 0.361, SE = 0.051; \text{Mean Difference} = 0.106, 95\% \text{ CI [0.000, 0.212]}, p<0.05$) and DLPFC stimulation marginally impaired JOL accuracy ($M = 0.164, SE = 0.050; \text{Mean Difference} = -0.091, 95\% \text{ CI [-0.194, -0.11]}, p<0.09$). This produced a significant difference in JOL accuracy for inverted words between DLPFC and aPFC stimulation sites ($\text{Mean Difference} = 0.197, 95\% \text{ CI [0.096, 0.299]}, p<0.001$) (Fig. 5B). Similarly to the effects of stimulation on $d_a$ for high- and low-frequency words, here, aPFC stimulation appears to have selectively improved JOL accuracy for words in the less fluent condition.

**Discussion**

Although the aPFC has been implicated in JOL accuracy (Kao et al., 2005; Ryals et al., 2016), and the DLPFC has been associated with JOL magnitude and encoding success (Blumenfeld & Ranganath, 2007; Do Lam et al., 2012; Kao et al., 2005), no research to date has investigated how the roles of these regions may vary with the cognitive bases of JOLs. We replicated previous behavioral findings by showing that mean JOL ratings were higher for high-frequency and upright words, consistent with an ‘easily learned, easily remembered’ heuristic (Koriat, 1997). However, because memory performance benefits from depth of encoding (Bjork & Bjork, 2011; Craik & Lockhart, 1972) and stimulus novelty (Kishiyama, Yonelinas, & Knight, 2009), recognition was better for low-frequency and inverted words. Turning to the brain bases of JOLs and encoding, HD-tDCS over the DLPFC impaired encoding, as evidence by increased false alarms for all cue types. For JOL accuracy, HD-tDCS over the aPFC
selectively improved JOL accuracy for the least fluent encoding conditions, i.e. low-frequency and inverted words. Taken together, our results suggest a causal role of the DLPFC in encoding, indicate that memory and metamemory functions are at least partially dissociable in the prefrontal cortex. and fill an existing gap in knowledge about the interaction between cognitive and neural bases of JOLs by revealing that the roles of the aPFC and DLPFC in metamemory accuracy vary based on encoding fluency.

Behavioral Effects of cue type on memory and metamemory

Our findings that word frequency and orientation predicted recognition performance are consistent with past research suggesting subjects have better memory under conditions of encoding disfluency (Besken & Mulligan, 2013; Yue et al., 2012), including disfluency that results from lack of familiarity with the stimulus (Jia et al., 2016; Kishiyama & Yonelinas, 2003; Whittlesea & Williams, 2000). Several studies have shown that the novelty of verbal stimuli influences recognition success (Kishiyama & Yonelinas, 2003; Kishiyama et al., 2009), likely because novelty leads to deeper processing or attentional orienting to unfamiliar stimuli (Corbetta & Shulman, 2002; Kishiyama et al., 2009). In the current study, subjects had higher hit rates and lower false alarm rates for low-frequency as compared to high-frequency words, confirming the benefit of relative novelty on recognition performance. We also found effects of word orientation on memory performance: subjects had significantly higher hit rates and lower false alarm rates for inverted as compared to upright words, consistent with past research suggesting perceptual disfluency promotes memory success due to deeper encoding (Besken & Mulligan, 2013; Rhodes & Castel, 2008).

Despite having better overall memory for low-frequency and inverted words, JOLs were significantly higher for high-frequency and upright words. This is consistent with past research showing the effects of word frequency on metamemory ratings (Benjamin, 2003; Jia et al., 2016; Sungkhasetee et al., 2011), and supports the notion that subjects use an ‘easily learned, easily remembered’ heuristic wherein confidence increases with fluency of encoding (Besken & Mulligan, 2013; Koriat, 2008a; Miele, Finn, & Molden, 2011). Because low-frequency words are more disfluent due to less baseline familiarity with the stimulus (Balota, Burgess, Cortese, & Adams, 2002; Jia et al., 2016), and processing of inverted words is more effortful due to perceptual disfluency (Johnston et al., 1985; Rhodes & Castel, 2008),
subjects rely on fluency as a cue to learning and mistakenly give lower JOLs to words under these conditions, based on the subjective experience of more effortful encoding.

**Effects of HD-tDCS on recognition performance**

In addition to the behavioral effects of frequency and orientation on recognition performance, we showed that HD-tDCS over the DLPFC led to greater subsequent false alarms, as compared to aPFC and sham stimulation. This pattern held for the model that included frequency, with greater false alarm rates for both high and low frequency words, and for the model that included orientation, with greater false alarm rates for both upright and inverted words. These results suggest the DLPFC plays a causal role in encoding success by supporting accurate discrimination between old and new items, and stimulation interfered with this function.

Most past literature implicating the DLPFC in false alarms has focused on its contribution during retrieval (Henson et al., 1999; Parkin, Bindschaedler, Harsent, & Metzler, 1996; Yonelinas, Otten, Shaw, & Rugg, 2005), but some research has shown the role of the lateral PFC during encoding is associated with subsequent false alarms (Demeter, Mirdamadi, Meehan, & Taylor, 2016; Slotnick & Schacter, 2004). For example, one fMRI study showed that increased activity in the lateral PFC (BA10/45) during encoding of shapes was associated with subsequent false-alarms to similar nonstudied shapes as compared to subsequent hits (Slotnick & Schacter, 2004), although it is worth noting that these regions are more ventral/anterior than our DLPFC stimulation site. However, another study showed that excitatory short theta-burst stimulation (TBS) to the left DLPFC during encoding marginally lowered the proportion of subsequent false alarms for lure items relative to stimulating the vertex (Demeter et al., 2016), indicating perhaps short TBS to the DLPFC strengthened encoding of specific information, which lead to a decrease in false alarms.

False alarms in recognition memory are typically thought to be based on familiarity (Kishiyama & Yonelinas, 2003; Reder et al., 2000; Yonelinas, 1994), a hypothesis which is further supported by our finding that subjects had greater false alarm rates for high-frequency than low-frequency words. In addition to feelings of familiarity that stem from repeated exposure to a stimulus, such as in the case of high-frequency words, there is also evidence that perceptual fluency may lead to a sense of familiarity
(Johnston et al., 1985; Proverbio et al., 2007), supporting our finding that subjects had higher false alarms for upright as compared to inverted words. Although the effects of DLPFC stimulation on false alarms did not vary by cue type in the current study, other work has implicated the DLPFC in encoding of more distinctive items. Indeed, several lesion studies have demonstrated that patients with DLPFC lesions do not show the memory advantage for novel stimuli typically seen in healthy controls (Duarte et al., 2005; Kishiyama & Yonelinas, 2003; Kishiyama et al., 2009). In one study, prefrontal lesion patients completed a von Restorff task, in which photos of items were made novel at study by presenting them in a color that differed from the majority of other study items. Subjects had poorer recollection- and familiarity-based recognition performance for all items compared to controls and, importantly, did not show a memory benefit for novel as compared to familiar photos, suggesting the DLPFC is crucial to the novelty advantage in long-term memory encoding (Kishiyama et al., 2009). Another lesion study showed that patients with DLPFC lesions were impaired in familiarity-based recognition, and that this impairment was specific to items presented toward the lesioned hemisphere at encoding (Duarte et al., 2005), suggesting a deficit in successful subsequent familiarity processing at encoding, rather than retrieval. Therefore, one possible interpretation of the effect we found of DLPFC stimulation on false alarms is that recruitment of the DLPFC during encoding differs for novel vs. familiar stimuli, and tDCS to this region disrupted the benefit of stimulus novelty on subsequent performance.

It is worth noting that our results show that tDCS to the DLPFC during encoding disrupted, rather than enhanced, recognition performance by increasing false alarms. Although some studies have shown that tDCS to the prefrontal cortex can benefit memory performance (Chua et al., 2017; Matzen, Trumbo, Leach, & Leshikar, 2015), several other studies have shown that stimulation fails to improve performance in other cognitive domains (Boggio et al., 2010; Gaynor & Chua, 2016; Marshall, Mölle, Siebner, & Born, 2005; Monti et al., 2008), and this is likely due to differences in task design, stimulation parameters such as electrode montages, and characteristics of the subjects in the sample (Tremblay et al., 2014). Therefore, our finding that tDCS to the DLPFC impaired recognition performance is consistent with research suggesting that excitatory stimulation does not always have facilitatory effects on performance (Boggio et al., 2010; Gaynor & Chua, 2016; Marshall et al., 2005; Monti et al., 2008; Tremblay et al., 2014), and demonstrates a causal role of the DLPFC in encoding success.
Effects of HD-tDCS on JOL accuracy

Based on past research implicating the aPFC in JOL accuracy (Kao et al., 2005; Ryals et al., 2016), and evidence that the aPFC plays a role in integrating information from other regions in support of abstract rules and goals (Badre, 2008; Koechlin et al., 1999), we hypothesized that aPFC stimulation would bias individuals toward using cues that are predictive of memory success, leading to better JOL accuracy in the low frequency and inverted conditions. Indeed, aPFC stimulation significantly improved JOL accuracy, as measured by $d_a$, for low-frequency and inverted words, relative to sham and DLPFC stimulation, showing that the aPFC plays a causal role in JOL accuracy. However, we also found that both DLPFC and aPFC stimulation decreased JOL accuracy for high-frequency words relative to sham, suggesting the roles of these regions vary based on the diagnosticity of the cue: for disfluent conditions, the aPFC made subjects’ JOLs more sensitive to the memory benefit of disfluency, but in the case of high-frequency words, which are familiar and fluently encoded, DLPFC and aPFC stimulation made subjects more reliant on the nondiagnostic cue of fluency. The finding that the aPFC biased individuals toward using disfluency as a cue that predicts memory success is in line with the theory that when faced with multiple possible responses based on bottom-up information from posterior regions, anterior portions of the PFC act to bias responses in the interest of higher-order ongoing task goals (Badre, 2008). However, aPFC stimulation also made subjects’ JOLs less sensitive to the fact that high fluency of encoding is non-predictive of memory success. In other words, in the case of high-frequency words, subjects made JOLs that were even more dependent on fluency, resulting in poorer JOL accuracy. Therefore, it remains unclear how the aPFC may bias subjects toward using predictive information in the face of competing response options, because enhancement of aPFC activity increased reliance on both diagnostic and nondiagnostic cues. Nevertheless, the finding that aPFC stimulation influenced JOL accuracy is somewhat consistent with theories suggesting cognitive control mechanisms may be arranged hierarchically along the rostro-caudal axis of the PFC, with anterior regions supporting complex, abstract relationships between representations, such as the relationship between JOL ratings and encoding success (Badre, 2008; Christoff & Gabrieli, 2000).

In the context of theories proposing a hierarchical organization of the PFC, the finding that the
DLPFC also impaired JOL accuracy for high-frequency words could also reflect increased reliance on stimulus-level properties when making JOLs: if posterior regions of the PFC track lower-level sensory information related to the stimulus, enhancing activity in the DLPFC may have enhanced the salience of the fluency cue, making it a more prominent basis on which subjects based JOLs, resulting in poorer JOL accuracy. However, we did not find that DLPFC stimulation altered mean JOL ratings for any cue types, which might have been expected if stimulation enhanced the salience of cues on which JOLs are based, and further research should address whether the DLPFC plays a direct role in JOL accuracy, or whether it tracks only stimulus-level properties.

Limitations

Our inferences about the roles of the aPFC and DLPFC in JOLs and encoding rest on the assumption that we are administering focal stimulation to these precise regions using HD-tDCS. However, despite the ability of HD-tDCS to administer relatively more focal stimulation as compared to conventional tDCS (Villamar et al., 2013), it is still possible that current reached cortical regions outside the regions of interest in our study (Bai, Dokos, Ho, & Loo, 2014; Bikson et al., 2012; Datta et al., 2009), and that the differences we saw between DLPFC and aPFC stimulation conditions reflect relative differences between potentially broader regions stimulated by our montages. Indeed, multimodal studies have shown that stimulation may alter activity in more distal cortical regions that are functionally connected to the sites of stimulation (Hampstead et al., 2014; Jacobson et al., 2012; Keeser et al., 2011). Nevertheless, studies have shown that HD-tDCS stimulation is relatively more focal than conventional tDCS (Kuo et al., 2013; Villamar et al., 2013), and here we show relative differences between the effects of aPFC and DLPFC stimulation, suggesting that the roles of these regions in JOL processes are at least somewhat dissociable.

It is also important to consider that the effects of tDCS on behavior are known to be dependent on a variety of factors related to endogenous brain activity (Berryhill et al., 2014; Brunoni et al., 2012; Datta et al., 2012), which may vary based on anatomical differences, age, mood state, and levels of arousal. Therefore, it is possible that we failed to detect some effects of stimulation due to variability in responses to tDCS based on these factors. We attempted to minimize the degree to which these factors influenced
our results by controlling for individual differences in sex, age, head circumference, stimulation experience (e.g., itching, headache), and mood states (e.g., trouble concentrating, sleepiness) in our analyses of the effects of tDCS on memory and metamemory; however, further research is needed to test how these individual differences mediate the effects of tDCS on other forms of cognition, with an aim to better predict who will benefit from tDCS.

Conclusions

The current experiment provides evidence for a causal role of the DLPFC in recognition based on distinctiveness at encoding, and of the aPFC in JOL accuracy. Importantly, this is the first study to investigate how the cognitive basis of JOLs interacts with the regions of the PFC thought to be involved in JOL processes. We showed that the aPFC contributes to accurate JOLs made based on word frequency by biasing individuals toward basing their subjective judgments on disfluency, which is predictive of objective memory success. Conversely, DLPFC and aPFC stimulation decreased JOL accuracy for high-frequency words, suggesting the roles of these regions in JOLs vary with the cognitive bases of the judgments. We also demonstrated that HD-tDCS may be an effective method by which to enhance metamemory accuracy, but that it may also impair accuracy, and the nature of the effect varies with the cognitive basis of subjective confidence. The results of the current experiment provide an important contribution to our understanding of the causal roles of PFC subregions in memory and metamemory processes and how they may vary with cognitive mechanisms underlying these functions, and our findings have important clinical implications for the use of non-invasive brain stimulation to improve metamemory deficits.
Chapter 3

Brain activity associated with confidence in semantic recognition varies based on the presence and validity of external cues
Abstract

Past research has shown that retrospective confidence judgments (RCJs), in which subjects assess their confidence in a retrieval response, can be based on multiple sources of information, including the strength of the memory trace, internal cues stemming from the subjective experience of retrieval, and external information about the task. Studies on the neural mechanisms involved in RCJs have implicated the prefrontal, parietal, and temporal cortices in the magnitude and accuracy of these judgments in episodic retrieval, but no studies to date have tested how the roles of these regions may differ when RCJs are based on internal memory cues as compared to external questions about task difficulty during semantic recognition. Here we used fMRI to test brain activity associated with semantic recognition and RCJs when subjects received external information about task difficulty, in the form of valid, invalid, or no cues about question difficulty. We first replicated and extended prior work on RCJs in episodic memory tasks, and showed that, like in episodic tasks, making a confidence judgment compared to making a recognition judgment engaged regions of the prefrontal, parietal, and temporal regions during a semantic memory task. Next, focusing on the influence of external information, there was greater activity during uncued trials compared to trials with a difficulty cue in lateral and medial parietal regions, as well as inferior temporal regions, which may be related to response uncertainty. We found that activity in the parietal lobe, anterior cingulate, and occipital cortex was greater for valid than invalid cues during hard questions, whereas activity in occipital and subcortical regions was greater for invalid than valid cues during easy questions, suggesting the effects of cue validity vary with question difficulty. Finally, we examined individual differences and showed that the degree to which subjects relied on an invalid cue during RCJs for correctly answered hard questions correlated with activity in bilateral medial temporal lobes, which may reflect a role of this region in integration between multiple sources of information. Taken together, our results contribute to our understanding of the brain regions involved in RCJs during semantic retrieval, and suggest the roles of these regions differ when subjects make judgments based on internal retrieval-based signals as compared to external sources of information.
**Introduction**

Accurate confidence in one’s own memory is crucial to effective learning and behavioral control. For instance, a student who is confident he has retrieved the correct answer on a given exam question is less likely to spend further time on that question or check his answers, while a student with less confidence in his response may spend more time attempting retrieval of possible alternative answers.

Retrospective confidence judgments (RCJs) are subjective judgments made after retrieval, and reflect an individual’s certainty in the accuracy of their response (Nelson & Naren, 1990). Research on the cognitive bases of RCJs has shown that individuals make inferences based on a variety of information available at retrieval when making their confidence judgments, including: the strength of the memory being assessed, internal cues that arise from the experience of retrieval, and/or external information about task difficulty or preconceived notions about competence (Busey et al., 2000; Chua et al., 2012; Koriat & Levy-Sadot, 1999; Koriat et al., 2008; Norman & Schacter, 1997). However, less is known about the neural basis of confidence judgments based on different sources of information. Neuroimaging studies have implicated the prefrontal cortex (PFC) and posterior parietal cortex (PPC) in RCJs (Chua et al., 2006, 2009a, 2009b; Hayes et al., 2011; Henson et al., 2000; Kim & Cabeza, 2009; Simons et al., 2010). Some regions, such as the dorsolateral and ventrolateral PFC and dorsal parietal cortex, show greater activity with lower confidence responses, while others, such as the ventral parietal cortex, show greater activity with high confidence. However, there is some evidence that the roles of the PPC and PFC may vary with the cognitive basis of the RCJ (Kim & Cabeza, 2007). Here, we manipulated the validity of external information about question difficulty in a semantic recognition task to determine how frontal and parietal regions contribute to RCJs that are based on external information vs. internal cues about memory success.

Early research on the cognitive bases of RCJs promoted a direct access view of metamemory, based on the hypothesis that individuals monitor and assess the strength of stored memories, and base their confidence on the product of this assessment (Busey et al., 2000; Hart, 1967a). This hypothesis was supported by evidence that subjective confidence judgments tend to be relatively accurate as they relate to objective memory performance, and both processes are often sensitive to the same experimental manipulations (Cohen, Sandier, & Keglevich, 1991; Lindsay et al., 1998; Stretch & Wixted,
However, other work challenged the direct access views because the same experimental manipulations can have different effects on memory accuracy and confidence ratings (Jameson, Narens, Goldfarb, & Nelson, 1990; Kelley & Lindsay, 1993), which suggested confidence may be based on information other than, or in addition to, access to the strength of stored memories, and that metamemory and memory processes are dissociable.

Instead of direct access, a significant body of behavioral research has suggested that RCJs are inferential in nature, and the level of confidence expressed can be based on internal and external information available at the time of the judgment (Bradfield et al., 2002; Jaeger, Cox, & Dobbins, 2012; Jaeger, Lauris, et al., 2012; Koriat & Levy-Sadot, 1999; Koriat et al., 2008). For instance, a subject's confidence may be inflated because she quickly retrieved information during memory search in an attempt to answer a question (Koriat et al., 2008), and because accurate answers are often quickly retrieved, she infers this information is correct even if it was incorrect. In another example, her answer may be consistent with external sources (Jaeger, Lauris, et al., 2012), and she infers her response is correct even if those sources are unreliable. When RCJs are based on information that is correlated with actual memory success, there is a strong relationship between confidence and memory accuracy; however, confidence may also be based on cues that are nondiagnostic of actual memory performance, leading to inaccurate RCJs (Busey et al., 2000; Koriat et al., 2008). Thus, it is critical to understand under what conditions subjects base their confidence on different types of information available at retrieval.

The dual-process system of metacognitive monitoring posits that individuals base their confidence judgments on: 1) experience-based cues, which arise from the subjective experience of retrieval, and 2) information-based cues, those that reflect previously held beliefs about one's memory, test difficulty, and competence in the domain being tested (Koriat et al., 2008). Common experience-based cues are typically related to memory strength, such as vividness of the accessed memory (Robinson, Johnson, & Robertson, 2000), which lead to relatively accurate confidence ratings. However, RCJs are also sensitive to other experience-based cues that are not necessarily diagnostic of memory accuracy, such as retrieval fluency (Shaw & McClure, 1996). For instance, in weeks following witnessing an event, subjects had inflated confidence in their memory for details about which they are repeatedly tested, relative to those details about which they were asked only at initial questioning, but had equal
memory accuracy for both conditions, suggesting that repeated retrieval lead to a greater sense of fluency on which confidence assessments were based (Shaw & McClure, 1996).

Information-based cues, which reflect beliefs about one’s own memory processes or factors that affect memory performance, have also been shown to influence confidence judgments (Jaeger, Lauris, et al., 2012; Koriat et al., 2008; Wells & Bradfield, 1998, 1999). For instance, when subjects were asked to list reasons in support of their answer on a general knowledge recognition task, confidence increased with the number of reasons listed, perhaps because subjects weighed the evidence in favor of their answer relative to evidence in favor of the alternative, and thus had higher confidence when they were able to retrieve more supportive evidence (Koriat et al., 2008). Another source of external information that influences confidence judgments is “consensuality”, or the proportion of people who endorse a given response, regardless of its accuracy (Koriat et al., 2008). For instance, Koriat et al. (2008) found that when subjects studied English words and their translations in a foreign language, confidence ratings correlated with the consensuality of the translation: subjects had higher confidence for both correct and incorrect answers when those answers were endorsed by most other participants (Koriat et al., 2008). This suggests that individuals inflate their confidence for questions to which they believe most people should know the answer, irrespective of the actual accuracy of the response.

Although cues can be categorized as experience or information based, individuals often use a combination of information-based and experience-based cues when making confidence judgments, and the weights given to different type cues vary based on the individual and situation. For example, when individuals have weak memories they may rely less on internal experience-based cues, and more heavily on external information-based cues, such as recommendations of confederates, (Bradfield et al., 2002; Jaeger, Lauris, et al., 2012; Wells & Bradfield, 1998, 1999). Specifically, eyewitnesses’ confidence in their identification of targets in a lineup is inflated when they receive feedback confirming their choice, and this may be because of weak memory of the conditions of retrieval, such as having a clear view of the suspect (Wells & Bradfield, 1998, 1999). Indeed, one study testing this hypothesis found that inaccurate witnesses’ confidence was higher when receiving confirming feedback than in a control condition, but accurate witnesses gave equal confidence ratings with and without feedback, indicating those with weaker internal memory cues are more affected by external feedback (Bradfield et al., 2002). Similar
results, in which subjects incorporate external information into their confidence ratings when unable to rely on internal memory cues, has been shown in old/new recognition tasks (Jaeger, Cox, et al., 2012; Jaeger, Lauris, et al., 2012). In one study, subjects were given cues about whether an item was ‘likely old’ or ‘likely new’, which were either from reliable (75% valid) or unreliable (50% valid) sources. In cases of low confidence, subjects were more likely to incorporate the recommendations of both reliable and unreliable sources, likely because they could not rely on their own internal memory cues during such trials (Jaeger, Lauris, et al., 2012).

Although behavioral research has addressed some of the cognitive bases of RCJs, relatively less is known about the brain bases of these judgments, particularly how brain activity associated with RCJs relates to the cognitive bases of the judgments (Chua et al., 2014; Kim & Cabeza, 2007). There is some initial evidence that the roles of brain regions involved in RCJs may vary with the cognitive sources of confidence (Kim & Cabeza, 2007), but a better understanding is needed of the neurocognitive mechanisms involved in the reliance on internal vs. external sources of information when making confidence judgments, which is the goal of this study.

There are multiple ways to examine confidence-related processes in the brain, including comparing the brain activity correlated with the process of assessing one’s confidence as compared to the process of retrieval, and which regions track the subjective level of confidence expressed (Chua et al., 2006). Regions of the prefrontal and parietal cortices have been broadly implicated in RCJ processes (Chiu et al., 2011; Chua et al., 2009a; Davidson et al., 2008; Hayes et al., 2011; Henson et al., 2000; Kim & Cabeza, 2007; Simons et al., 2010). Lesion studies have shown that RCJs are impaired in frontal traumatic brain injury (TBI) patients in a recognition task (Chiu et al., 2011), and in patients with parietal lesions in remember/know paradigms (Davidson et al., 2008), and old/new recognition and source recollection tasks (Simons et al., 2010). Furthermore, neuroimaging research has shown that the process of making a confidence judgment, as compared to a recognition task, is associated with activity in the lateral and medial parietal cortex, and dorsolateral, ventrolateral, and ventromedial prefrontal cortex (Chua et al., 2006, 2009b).

Turning to the subjective level of confidence expressed, neuroimaging data has shown that in the prefrontal cortex, the DLPFC and ventrolateral PFC show more activity during low as compared to high
RCJs in an associative recognition paradigm (Chua et al., 2009a), an old/new verbal recognition task (Henson et al., 2000), and during item and source memory tasks (Hayes et al., 2011), perhaps because lower confidence trials place greater demands on monitoring processes. In the parietal cortex, activity in the dorsal PPC was greater for low vs. high confidence (Chua et al., 2009a; Kim & Cabeza, 2007; Moritz et al., 2006), consistent with research suggesting the superior parietal lobe is involved in top-down allocation of attention to memory retrieval (Ciaramelli et al., 2008; Olson & Berryhill, 2009) and increased activity when monitoring uncertain or ambiguous targets that require more effortful retrieval (Ciaramelli et al., 2008).

In contrast to dorsal PPC regions that show increased activity with low confidence, studies have shown that activity in the ventral PPC correlated with higher confidence ratings (Chua et al., 2009a; Hayes et al., 2011). This is consistent with research suggesting the ventral PPC may track the strength of the memory being assessed (Cabeza et al., 2008), as evidenced by greater activity during recollection than familiarity-based recognition (Henson et al., 1999), and during hits as compared to correct rejections (Konishi et al., 2000). Taken together, these results suggest PFC and dorsal PPC regions show greater activity with low confidence due to greater demands on monitoring during effortful retrieval, while activity in the ventral PPC tracks the strength of the memory being monitored.

Although the above research has begun to identify how prefrontal and parietal regions contribute to high and low confidence judgments, it remains unknown how the neural mechanisms underlying RCJs interact with the cue types on which RCJs are based. However, there is some evidence to suggest that the roles of prefrontal and parietal regions in RCJs may vary with the cognitive basis of the judgments (Kim & Cabeza, 2007). For example, Kim & Cabeza (2007) showed greater activity in frontal and lateral parietal regions with low confidence during true recognition, but greater activity in similar regions with high confidence during false recognition. Based on the premise that false recognition is likely based on a sense of familiarity, while true recognition is based on recollection, activity in the PFC and parietal cortex appears to be sensitive to the underlying cognitive basis of the confidence judgment (Kim & Cabeza, 2007). Work by O’Connor, Han, & Dobbins (2010), also informs the relationship between brain activity and the cognitive bases of recognition judgments; when subjects were given cues from reliable and unreliable sources about the likelihood a target was old or new during an episodic recognition task, there
was greater fMRI activity in the inferior parietal lobule and the medial PFC with invalid as compared to valid cueing (O’Connor et al., 2010), which may reflect the recruitment of cognitive control processes that are engaged when internal memory signals and external information-based cues conflict.

Much of the research to date has examined the neurocognitive underpinnings of RCJs in episodic memory tasks (Chua et al., 2009b, 2009a; Hayes et al., 2011; Henson et al., 2000), and fewer studies have examined the neural correlates of RCJs in a semantic retrieval task. Furthermore, the small number of studies that have manipulated the validity of external cues in RCJs have used episodic recognition tasks (Jaeger, Cox, et al., 2012; Jaeger, Lauris, et al., 2012), leading to the question of whether the effect of cue validity on confidence-related activity is similar in semantic retrieval. Past research has indicated that although there is some overlap between regions that contribute to episodic and semantic retrieval (Burianova & Grady, 2007), episodic and semantic retrieval tend to recruit different brain regions; namely, semantic retrieval frequently engages the left prefrontal cortex, while episodic retrieval engages right prefrontal regions (Nyberg, Cabeza, & Tulving, 1996). However, there is evidence from other metamemory tasks that metamemory judgments during semantic and episodic retrieval recruit some overlapping brain regions (Reggev, Zuckerman, & Maril, 2011). Thus, an aim of the current study is to address the gap in knowledge regarding how cue validity effects RCJs in semantic retrieval, and how this compares to previous work that has identified regions that contribute to confidence in episodic retrieval.

The current study examines how brain activity varies when confidence judgments are based on different sources of information in a semantic retrieval task, with a specific interest in internal experience-based versus external information-based cues (Bradfield et al., 2002; Jaeger, Lauris, et al., 2012; Kroriat et al., 2008). Given that the parietal and prefrontal cortices are involved in retrospective confidence judgments (Chua et al., 2006, 2009b, 2009a; Hayes et al., 2011; Henson et al., 2000; Simons et al., 2010), we expected that activity in the PFC and PPC would modulate depending on the information on which confidence is based. In our task, subjects completed a recognition and RCJ task in which they chose one of four possible answers to general knowledge questions, and then gave a confidence rating on a scale of 1-4 for each answer. Two-thirds of trials included a cue about the difficulty of the question (‘Easy’ or ‘Hard’) and one-third were uncued; cues were valid 50% of the time. This allowed us to examine brain activity for invalid vs. valid cues, cued trials vs. uncued trials, and how these differences
varied by actual question difficulty, as well as differences in activity during the recognition as compared to confidence tasks, with an aim to better understand the neurocognitive correlates of RCJs based on experience-based and information-based cues.

**Methods**

**Participants**

Twenty healthy adults were consented to participate for financial compensation ($25/hour for 2.5 hours). Two participants were excluded due to signal instabilities during acquisition. Data from 18 participants (12 F, ages 18-35, M = 22.8, SD = 3.7 years) were included. All participants were right-handed, learned English before the age of 5, and were free from any self-reported neurological or psychological disorders, chronic illnesses, or contraindications for fMRI (e.g., no metallic implants, claustrophobia). All participants gave written consent in a manner approved by the Human Research Protection Program at the City University of New York.

**Behavioral Task**

Participants were given a four alternative forced choice recognition test for 300 general knowledge questions from the Baruch Knowledge Norms (http://www.mangelslab.org/bknorms), which consists of a database of 406 questions related to natural and social science, mathematics, technology, history, geography, arts and culture. Questions were divided by difficulty (Easy and Hard) using a median split in performance based on data collected from 498 CUNY students who had previously answered the same questions. Each question was presented for 7s, with 4 possible answer choices, including one correct choice that was unique to one question, and 3 distractors consisting of the most commonly given incorrectly recalled answers. Subjects indicated their response choice using the numbers 1-4 on the MRI button box. The key press position of the correct answer (i.e., key press 1-4) was counterbalanced across trials. For 500ms prior to, and throughout the duration of the presentation of the question (for a total of 7.5s), participants saw a difficulty cue at the top of the screen. They were instructed that an “Easy” cue indicated that most people tended to answer that question correctly, even if they felt they were guessing, “Hard” indicated that most people answered incorrectly, and “????” indicated there was no past
performance data for that particular question. Prior to the start of the fMRI task, participants were given a practice task outside the scanner, which included feedback on accuracy of their responses to 24 sample questions (12 hard, 12 easy questions; one third with “Easy” cue, one third with “Hard” cue, one third uncued). For all questions cued as “Easy”, participants received feedback that their answer was correct, regardless of actual accuracy. For “Hard” cues and Uncued questions, participants received accurate performance feedback. This was done to make the instructions stating that cues were valid and based on typical subject performance more believable. During the actual task in the scanner, “Easy” and “Hard” cues were valid 50% of the time, and the true difficulty of Uncued questions was 50% Easy and 50% Hard. Each recognition response was followed by presentation of a confidence rating scale (2.5s), during which subjects rated their confidence in having correctly answered the previous question on a scale from 1-4, in which a response of 1 represented 0-25% confidence, 2 = 25-50% confidence, 3 = 50-75% confidence, and 4 = 75-100% confidence (Fig. 6). Recognition and Confidence trials were intermixed with jittered interstimulus intervals consisting of periods of visual fixation ranging from 0-15s (M = 2.55, SD = 3.50), and optimized using Optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq). Total task time was 75 minutes, divided into 6 runs of 12.5 min each.

Behavioral Performance: Data Analyses

We used repeated-measures ANOVAs and post-hoc t-tests in SPSS 23.0 to test the effects of question difficulty and cue type on mean recognition performance, and the effects of question difficulty, cue type, and recognition performance on mean retrospective confidence ratings. Because we expected participants would incorporate the cues into their confidence judgments to different degrees, we created cue dependency scores for each subject, which reflected the degree to which their confidence ratings were influenced by a given cue type under each condition of question difficulty and memory performance, relative to an uncued condition. For example, a subject’s cue dependency score for invalidly cued hard questions would be calculated by subtracting the mean confidence rating for Uncued Hard questions from the mean confidence for Invalidly Cued Hard questions. Therefore, positive scores indicate the subject’s confidence was inflated by the presence of the cue, negative scores indicate confidence was lower with the cued relative to uncued condition, and a score of 0 would mean the cue had no impact on the
subjects’ confidence. This provided an individual measure of cue dependency for each subject, for each condition of cue type, question difficulty, and recognition performance. These individual cue dependency scores were used as covariates in whole-brain fMRI analyses to determine how brain activity varied with individual differences in how subjects incorporated cues into their confidence ratings (see below).

MRI Data Acquisition

Data were acquired on a Siemens Allegra 3T head-only MRI scanner with a custom head coil (NM-011; Nova Medical) at the New York University Center for Brain Imaging (New York, NY). Structural images were acquired with a high-resolution T1-weighted whole-brain scan using a 3D MP-RAGE sequence (1mm isotropic resolution, Field of View = 176mm, 176 slices). BOLD contrast data were acquired with a gradient-echo echoplanar imaging pulse sequence (TR = 2500ms, TE = 15ms, Flip angle = 82°, Field of View = 126mm, 42 slices positioned parallel to the anterior commissure-posterior commissure line, voxel size 3x3x3 mm). Visual stimuli were displayed using an LCD video projector with a projection screen positioned behind the subject’s head, and viewed through a mirror. Responses were collected using a button box held in the right hand. Each functional run consisted of 306 time points and lasted 12 min and 45 s.

Data Preprocessing and Analyses

Reconstructed EPI images underwent standard preprocessing using SPM8 (Wellcome Trust Centre for Neuroimaging). The first four scans of each run were discarded, and the remaining scans were corrected for slice timing, motion corrected by realigning to the mean functional image, coregistered to the given subject’s anatomical image, normalized to the standard space image in Montreal Neurological Institute (MNI) space, and smoothed with a 6-mm full-width half-maximum Gaussian kernel.

Trial types were categorized by cue type and actual question difficulty, producing 6 trial types: Valid Easy (i.e., Easy question with “Easy” cue), Invalid Easy (i.e., Easy question with “Hard” cue), Uncued Easy (i.e., Easy question without cue), Valid Hard (i.e., Hard question with “Hard” cue), Invalid Hard (i.e., Hard question with “Easy” cue), and Uncued Hard (i.e., Hard question without cue).

To examine the relationships between question difficulty and cue type on brain activity, we
defined a first-level GLM based on trial types, with each of the 6 trial types further defined by the cognitive task within the whole trial (i.e., Recognition response and Confidence rating for each given trial), producing 12 conditions of interest (6 for Recognition and 6 for Confidence Rating). To account for variance between runs, trials from each run were in separate regressors. The model used the canonical hemodynamic response function and included the durations of tasks (Recognition: 7.5s; Confidence: 2.5s). Individual run-specific motion regressors were also included in the model. First-level analyses produced individual maps of parameter estimates for all contrasts of interest, which were then submitted to a second level group analysis. Timecourses of hemodynamic response function were modeled using a finite impulse response function using MarsBar toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002). Contrasts of activity during Confidence vs. Recognition trials were corrected for multiple comparisons using family-wise error correction (p = 0.05). Due to low statistical power and the exploratory nature of analyses testing the effects of the presence and validity of cues on brain activity, all other analyses were uncorrected for multiple comparisons. All results were considered significant at p<0.001, with a voxel extent threshold of 10. Voxel coordinates are reported in MNI space.

SPM contrast maps were generated for the main contrasts of interest, including:

1) **Confidence task vs. Recognition task.** Significant differences in activation shown in this contrast represent brain regions that differentially contribute to each process.

2) **Cued vs. Uncued trials.** Comparisons of Cued vs. Uncued trials were assessed during recognition and confidence tasks separately. Significant differences in activation shown in these contrasts represent brain regions that differentially respond to the presence vs. absence of an external cue about question difficulty, tapping into the contribution of external information to recognition and confidence decisions.

3) **Valid vs. Invalid cued trials.** Comparisons of Valid vs. Invalid trials were assessed during recognition and confidence tasks separately. Significant differences in activation shown in these contrasts represent brain regions that differentially respond based on the validity of the external cue. Based on behavioral findings, fMRI activity during Valid vs. Invalid cues were assessed separately for Hard and Easy questions; this allowed us to examine how differences in activity based on cue validity may vary with question difficulty.
Individual Differences. To examine how confidence ratings and memory performance interacted with trial types, we used a cue dependency score, which reflected how much each individual’s confidence rating was altered by a given cue. Cue dependency was calculated by subtracting the mean confidence rating for an Uncued condition from the mean confidence for an associated cued condition, e.g., Invalidly Cued Hard Hits – Uncued Hard Hits. We conducted analyses using the individual cue dependency scores as covariates in SPM in the whole-brain analyses of Confidence > Recognition trials. These analyses examine how activity associated with making confidence judgments varied based on individual differences in cue usage. Mean percent signal change for all voxels within regions of interest (ROI) was extracted using MarsBar to illustrate how individual cue usage covaried with activity in each condition within the contrast of interest.

Results

Behavioral Results

Confidence Judgments. Mean confidence ratings across all question difficulty and cue conditions ranged from 1.79 to 3.35 (M = 2.47, SD = 0.40). We conducted a 2 (question difficulty: Easy, Hard) x 3 (cue type: Valid, Invalid, Uncued) x 2 (recognition performance: Hits, Misses) repeated measures ANOVA on mean confidence ratings. There was a significant main effect of question difficulty [F(1,17) = 189.867, p<0.001], with higher confidence ratings for easy questions (M = 2.606, SD = 0.076) than for hard questions (M = 2.134, SE = 0.087; Mean Difference = 0.427, 95% CI [0.399, 0.544]); a significant main effect of recognition performance [F(1,17) = 63.899, p<0.001], with higher confidence ratings for hits (M = 2.730, SE = 0.105) than misses (M = 2.010, SE = 0.075; Mean Difference = 0.720, 95% CI [0.530, 0.910]); and a significant main effect of cue type [F(2,34) = 4.421, p<0.05], with higher confidence ratings for questions with invalid cues (M = 2.426, SE = 0.089) compared to valid cues (M = 2.333, SE = 0.067; Mean Difference = 0.093, 95% CI [0.010, 0.176], p<0.05). Confidence for questions with invalid cues was also significantly higher than uncued questions (M = 2.350, SE = 0.088; Mean Difference = 0.076, 95% CI [0.021, 0.130], p<0.01), and there was no difference in confidence between uncued and validly cued questions (p=0.607). These main effects were qualified by a significant interaction between the effects of question difficulty and performance [F(1,17) = 59.812, p<0.001], and a marginal interaction between
question difficulty, cue type, and performance \([F(2,34) = 3.204, p<0.054]\) on mean confidence ratings. There was no significant interaction between question difficulty and cue type \((p>0.79)\) or cue type and performance \((p>0.87)\) on confidence.

Pairwise comparisons were conducted to follow up on the significant interaction between question difficulty and performance, and revealed that for hits, subjects had significantly higher confidence for easy questions \((M = 3.171, SE = 0.100)\) than for hard questions \((M = 2.289, SE = 0.120; \text{Mean Difference} = 0.883, 95\% \text{CI} [0.742, 1.023], p<0.001)\), but for misses, confidence for easy questions \((M = 2.040, SE = 0.093)\) did not differ from hard questions \((M = 1.980, SE = 0.065; p>0.32)\).

Although the question difficulty x cue type x performance interaction did not quite reach significance \((p<0.054)\), we used pairwise comparisons to examine differences in mean confidence ratings between these conditions. For easy questions, confidence did not differ between cue types for hits or misses, but for hard questions, there were two comparisons that indicated that participants were incorporating the cue in their confidence judgments. For hard questions, participants had marginally higher confidence for invalidly cued hits \((i.e., \text{Hard questions called "Easy" that they got correct;} M = 2.408, SE = 0.130)\) compared to validly cued hits \((i.e., \text{Hard questions called "Hard" that they got correct;} M = 2.248, SE = 0.126; p<0.096)\) and compared to uncued hits \((i.e., \text{Hard questions with no cue that they got correct;} M = 2.209, SE = 0.139; p<0.059)\). We calculated cue dependency scores \(\text{(see Behavioral Performance: Data Analyses, above)}\) to determine the degree to which individual subjects relied on the cue in making their confidence judgments, and found significant variability in these scores for Invalidly Cued Hard Hits \((\text{Fig. 7})\). Therefore, we used these individual cue dependency scores as covariates in fMRI data analyses to determine how brain activity was correlated with individual differences in cue dependency for Invalidly Cued Hard Hits.

**Recognition.** Mean recognition performance \((\text{proportion correct})\) across all conditions based on question difficulty and cue type ranged from 0.12 to 0.92 \((M = 0.48, SD = 0.22)\), and mean overall recognition performance ranged from 0.27 to 0.77 \((M = 0.48, SD = 0.13)\). We conducted a 2 \((\text{question difficulty: Easy, Hard})\) x 3 \((\text{cue type: Valid, Invalid, Uncued})\) repeated-measures ANOVA on mean recognition performance and showed a main effect of question difficulty \([F(1,17) = 127.787, p<0.001]\), no
effect of cue type (p>0.72), and a marginal interaction between question difficulty and cue type on mean performance [F(2,34) = 2.605, p<0.089]. As expected, subjects had better memory performance for easy questions (M = 0.640, SE = 0.040) than hard questions (M = 0.313, SE = 0.026; Mean Difference = 0.326, 95% CI [0.265, 0.387]). Although the question difficulty \times cue type did not quite reach significance, we followed it up with pairwise comparisons for Easy and Hard questions separately. There was no difference in recognition performance based on cue type for Easy questions. In contrast, for Hard questions, recognition performance was marginally higher for invalid cues (i.e., Hard questions called “Easy”; M =0.336, SE = 0.028) than valid cues (i.e., Hard questions called “Hard”; M = 0.302, SE = 0.026; p<0.053). Memory performance for uncued hard questions (M = 0.302, SE = 0.030) did not significantly differ from validly cued (p>0.99 or invalidly cued (p=0.142) hard questions. Therefore, if anything, any interaction between cue type and question difficulty on mean recognition performance was driven by better memory for invalidly cued hard questions as compared to validly cued hard questions, whereas performance did not vary by cue type for easy questions.

**Imaging Results**

**Confidence vs. Recognition Tasks**

Because a goal of the experiment was to examine activity associated with the process of making a confidence judgment, we first examined the contrast Confidence > Recognition, and showed greater activity for the confidence task in several large clusters that covered regions in the medial prefrontal, medial parietal, medial temporal lobe, and lateral tempo-parietal cortices (Table 1; Fig. 8). Activity in medial prefrontal, parietal, temporal, and lateral tempo-parietal regions is consistent with past research showing greater activation in similar regions during confidence ratings as compared to episodic recognition tasks (Chua, Rand-Giovannetti, Schacter, Albert, & Sperling, 2004; Chua et al., 2006, 2009b, 2009a), and these areas have been implicated in the default mode network, a network of regions that are thought to be activated during self-referential processing (Raichle et al., 2001).

We then examined activity associated with the process of recognition, by comparing activity during the recognition task to activity during the confidence task for all trials (Table 1). The contrast of Recognition > Confidence showed several large clusters that encompassed multiple regions (Table 1).
These clusters covered regions in left and right lateral prefrontal cortex (BA44), dorsal medial frontal cortex (BA6/8), the left parietal lobule (BA7/39), and visual cortices (BA17/18/19) (Fig. 9). More activity in these regions during recognition compared to confidence tasks is consistent with past research implicating the left ventrolateral prefrontal cortex (VLPFC) and lateral parietal cortex in semantic retrieval (Badre & Wagner, 2002; Wiggs, Weisberg, & Martin, 1998) (Table 1; Fig. 9).

**Cued vs. Uncued Trials**

A novel contribution of this study was to examine how brain activity differed based on the presence of an external cue, both during the confidence task and during the recognition task. During the confidence task, the contrast of Uncued > Cued trials revealed greater activity for Uncued trials in the right premotor cortex (BA6), inferior temporal lobe (BA37), left medial parietal lobe (BA23), left medial temporal lobe (BA36) and left cerebellum (Table 2). There were no significant clusters with greater activity for Cued than Uncued trials.

During recognition, the contrast of Uncued > Cued showed greater activity for Uncued trials in the left lateral prefrontal cortex (BA8), left inferior parietal regions (BA39), right lateral (BA7) and medial (BA31) parietal lobe, and right motor and somatosensory regions (BA4/1) (Table 2; Fig. 10). No regions showed significantly greater activity for Cued than Uncued trials during recognition.

**Valid vs. Invalid Trials**

In addition to the presence of a cue, it was also important to examine the effect of cue validity on brain activity during recognition and confidence tasks (Dobbins, Jaeger, Studer, & Simons, 2012; Jaeger, Cox, et al., 2012). During the confidence task, the comparison of activity during all Valid > Invalid cues showed greater activity for Valid cues in a small cluster (14 voxels) in the left visual association areas (-15, -58, 19; BA18). We then examined differences between valid and invalid cues for Easy and Hard questions separately, to test whether brain activity that differed based on cue validity varied with question difficulty. For hard questions, the Valid > Invalid contrast revealed greater activity centered in the angular gyrus (33, -76, 22; BA39) extending into visual cortices (BA17/18/19), the right anterior cingulate (24, 44, -5; BA32), right putamen (18, 14, 4), and left occipital gyrus (-30, -79, 13; BA19) during the confidence
task. For easy trials, the contrast of Invalid > Valid cues showed a small cluster in the right caudate (18, 26, -5) and in the right cuneus (18, -85, 4; BA18), suggesting the differences in activity for Valid and Invalid cues vary by question difficulty, with some regions showing more activity for invalid cueing during easy questions, and others showing increased activity for valid cues during hard questions.

During recognition, there were no significant clusters for the overall Invalid vs. Valid contrasts. For hard questions during recognition, there was greater activity for valid cues than invalid cues in the right parahippocampal gyrus (18, -10, -20). For easy questions, there were no significant differences in Valid > Invalid cues during recognition.

**Individual Differences in Cue Dependency**

Because there was significant individual variability in how subjects used cues about question difficulty to make confidence judgments (Fig. 7), we examined how significant activity in the Confidence > Recognition contrast, which is thought to represent brain areas engaged in making a confidence judgment, covaried with individual difference metrics of cue dependency. Based on behavioral results showing subjects had marginally higher confidence for Invalidly Cued Hard (i.e., Hard questions called “Easy”) questions answered correctly compared to Uncued Hard Hits (i.e., Hard questions with no cue), we included subjects’ cue dependency scores for the difference in confidence for Invalidly Cued Hard Hits - Uncued Hard Hits as covariates when comparing Confidence > Recognition. This allowed us to test how activity that was greater during Confidence than Recognition trials varied with the degree to which subject incorporated the invalid cue for hard hits into their confidence ratings.

There was a significant positive correlation between cue dependency scores and the difference in activity in Confidence > Recognition in bilateral medial temporal regions (L: -12, -7, -17; R: 15, -1, -20; Fig. 6). To investigate whether these differences were driven by activity in Confidence, Recognition, or both, we defined ROIs for the two MTL clusters showing a correlation, and extracted percent signal change for each condition. Bivariate correlations were used to illustrate the relationship between individuals’ dependency on an Invalid cue for a Hard question during a hit, and individual signal change during Recognition, and during Confidence. Individuals whose confidence ratings were more dependent on the Invalid cue for Hard hits had greater activity in left MTL (Fig. 11) and right MTL clusters during
confidence rating (i.e., the more an individual increased his or her confidence for a hard question that they answered correctly when it was labeled “Easy”, the greater the activity in the bilateral MTL during confidence rating). Correlation graphs also illustrated a negative relationship between cue dependency for Invalidly Cued Hard Hits and percent signal change during Recognition in the left MTL cluster, suggesting that the correlation between cue dependency and differences in activity between Confidence and Recognition was likely due to both an increase in activity during Confidence, and a decrease in activity during Recognition.

Discussion

We examined how brain activity varies when confidence judgments are based on different sources of information in a semantic retrieval task, with a focus on external information-based cues. We first showed that the brain regions with greater activity when making a confidence judgment compared to recognition in a semantic retrieval task, namely medial prefrontal, medial temporal, as well as medial and lateral parietal regions, were similar to those that have been previously identified as being involved in confidence in episodic tasks (Chua et al., 2006, 2009b). Furthermore, activity in a subset of these regions modulated based on the presence, validity, and/or usage of external information-based cues about question difficulty. Specifically, during the confidence task, brain activity in the medial temporal lobe and medial parietal cortex varied based on whether or not an external cue was given. Cue validity also played a role, and this varied based on task difficulty; for hard questions, there was greater activity in the lateral parietal cortex, centered in the angular gyrus for valid cues, but for easy questions, activity was greater in the caudate and cuneus for invalid cues. Finally, activity in the MTL covaried with how much individuals were influenced by the external cue. These findings, combined with previous fMRI (Chua et al., 2006, 2009a, 2009b; Hayes et al., 2011; Henson et al., 2000; Kim & Cabeza, 2007) and lesion studies (Chiou et al., 2011; Davidson et al., 2008; Simons et al., 2010) provide converging evidence for the roles of the PFC and parietal cortex in semantic recognition and subjective confidence judgments. Furthermore, in light of research showing individuals incorporate both internal and external cues into their RCJs (Bradfield et al., 2002; Jaeger, Cox, et al., 2012; Jaeger, Lauris, et al., 2012; Koriat, 2008b; Koriat et al., 2008), and the hypothesis that brain regions involved in RCJs may vary with the basis of the judgment (Kim &
Gaynor, 2007), our findings suggest activity in the parietal, prefrontal and medial temporal cortices modulates based on the presence and validity of external cues.

**The Process of Making Confidence Judgments & the Brain**

We were particularly interested in identifying regions engaged during the process of making a confidence judgment following *semantic recognition* because prior work has focused on confidence judgments in episodic tasks (Chua et al., 2006, 2009a). Similar to prior work in the episodic domain (Chua et al., 2006, 2009a), in the current study there was greater activity when participants engaged in making a RCJ compared to when they were making recognition judgment in the medial PFC (BA 9/10), medial parietal cortices, and lateral tempo-parietal cortices, suggesting common mechanisms underlying RCJs based on different types of recognition memory. The overlap between regions engaged in making a confidence judgment compared to recognition in the current and prior work (Chua et al., 2006, 2009a) is similar to other work in the domain of metamemory showing some overlap in brain activity associated with metamemory judgments for episodic and semantic memory (Reggev et al., 2011). In a feeling-of-knowing task, in which subjects judged their confidence in the ability to recognize an answer that was currently not retrievable, activity in the aPFC, middle frontal gyrus, dorsomedial PFC, anterior cingulate, and superior parietal lobe was associated with metamemory judgments irrespective of whether they were based on episodic or semantic retrieval (Reggev et al., 2011).

The set of brain regions that showed greater activity while making a confidence judgment compared to recognition resemble the “default mode network”, a collection of regions that are more active at rest, and thought to be involved in self-referential processing and introspection (Raichle et al., 2001). Specifically, activity in the dorsomedial PFC, one of the default mode regions with the greatest BOLD activity at rest, has been shown to increase during self-referential judgments (Davey, Pujol, & Harrison, 2016; Gusnard, Akbudak, Shulman, & Raichle, 2001; Mak et al., 2017; Meyer & Lieberman, 2018). Thus, one contribution of this study is that it replicates prior work identifying the brain regions involved in making confidence judgments in an episodic task (Chua et al., 2006, 2009a) and extends this work to a semantic task.

In addition to brain regions that showed greater activity during the RCJ task compared to the recognition task, there were also regions that showed greater activity during the recognition task
compared to the RCJ task, which were consistent with past fMRI research on both episodic and semantic retrieval (Badre & Wagner, 2002; Chua et al., 2006, 2009a). Specifically, there was greater activity in the medial and left lateral PFC when participants made recognition judgments compared to when they were making RCJs. Similar areas of the lateral and medial prefrontal cortex have been correlated with semantic retrieval (Badre & Wagner, 2002; Barredo, Verstynen, & Badre, 2016; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001; Wiggs et al., 1998). In particular, activity in the left medial PFC has been associated with semantic retrieval characterized by strong memory signals (Kim, 2016), and the lateral PFC has been identified as part of a control network involved in top-down control of memory retrieval (Barredo et al., 2016). Although past research directly comparing recognition to confidence assessment has focused on episodic memory tasks (Chua et al., 2006, 2009b, 2009a), some studies have shown overlap between regions implicated in both types of memory retrieval (Binder, Desai, Graves, & Conant, 2009; Burianova & Grady, 2007; Kim, 2016; Nyberg et al., 2003; Wiggs et al., 1998), supporting our finding that semantic recognition engages similar regions as episodic recognition, as compared to confidence rating.

**Information-based cues alter brain activity during recognition and confidence assessment**

A primary aim of the current study was to identify brain activity that varied with the presence and use of external cues about question difficulty during semantic retrieval. During the confidence task, there was greater activity for uncued than cued trials in the left medial parietal lobe (BA23), centered in the posterior cingulate, and also in the inferior temporal lobe (BA37), centered in the fusiform gyrus. Activity in the left medial parietal lobe/posterior cingulate is consistent with past research suggesting this region is recruited during evaluative judgments based on subjective values and preferences (Zysset, Huber, Ferstl, & von Cramon, 2002), and TMS research showing that disruption of the medial parietal cortex decreased episodic retrieval speed judgments made in reference to the self as compared to others, which suggests a causal role of this region in self-referential processing (Lou et al., 2004). Therefore, greater activity in the medial parietal cortex during uncued as compared to cued trials may reflect greater dependence on internal experience-based cues in the absence of external cues.
Although the recognition and confidence judgments were given as separate tasks, it is likely that some confidence-related processing occurred while making or even prior to making the confidence judgment (Chua et al., 2006), and so we also examined confidence-related activity during the recognition task. We showed that as compared to all cued trials, uncued trials showed greater activity during the recognition task in the left lateral PFC (BA8), right inferior parietal cortex (BA39), right lateral (BA7) and medial (BA7/31) parietal lobes, and right frontoparietal regions (BA4). The left lateral PFC has been broadly implicated in goal-directed cognitive control (Badre & Wagner, 2002; Nee & D'Esposito, 2016) and past research has shown greater activity in this region in response to both external uncertainty stemming from ambiguous circumstances, and internally attributed uncertainty, such as in the case of insufficient knowledge following memory search (Volz, Schubotz, & von Cramon, 2004). Thus, greater activity in this region for uncued as compared to cued trials in the current experiment may reflect response uncertainty resulting from a lack of external information about question difficulty, which places greater demands on memory search, or uncertainty due to the search producing inadequate information to make a decision. Uncued trials also showed greater activity in the medial parietal lobe during both the recognition and RCJ task, which may reflect a reliance on experience-based cues about memory success in the absence of an external cue. During memory search, retrieved information may elicit a feeling of confidence based on internal memory signals; if that confidence is insufficient to warrant reporting a response, however, memory search will continue until the subject can retrieve a response in which he/she is more confident. Thus, retrieval relies on a bidirectional relationship between memory search and monitoring processes. Therefore, during uncued trials, in which subjects may have to rely more on internal signals related to the experience of retrieval in order to assess their confidence and provide a recognition response, overlapping activity in the medial parietal cortex may reflect engagement of experience-based cues about memory success.

**Cue validity modulates the roles of parietal and temporal regions in RCJs**

In addition to examining how the presence of an external cue modulated the brain basis of confidence judgments, we also examined how the validity of the cue modulated the brain basis of confidence judgments. While participants were engaged in making confidence judgments, there was a
small cluster with greater activity for valid as compared to invalid cues in visual association areas, and no regions showing greater activity for invalid cues than valid cues. Turning to the effects of cue validity during the recognition task, again our results showed no regions with greater activity for invalid as compared to valid cueing, a result which differs from previous research during episodic retrieval showing that invalid cues during an old/new recognition task produced greater activity primarily the medial PFC and inferior parietal lobules (O’Connor et al., 2010). The difference between our results and O’Connor et al. (2010) suggests that the effects of cue validity on prefrontal and parietal activity vary by task type, such that the regions associated with invalid external cues during episodic recognition may differ from those associated with semantic recognition. However, it should be noted that when we reduced the significance threshold to p<0.01 for the Invalid > Valid contrast during recognition, there was significant activity in the medial PFC, suggesting the difference in results between our study and that of (O’Connor et al., 2010) may be due to lack of sufficient statistical power in the current experiment or the degree of differential activity in the medial PFC based on task type.

Whereas tasks in the episodic domain have relied on confidence judgments as a marker of difficulty (Chua et al., 2006, 2009a, 2009b; Dobbins et al., 2012; Hayes et al., 2011; Jaeger, Cox, et al., 2012), in the current study, we used an objective marker of general knowledge question difficulty, which allowed us to test how the use of external cues about difficulty interacts with internal experience-based cues that likely vary with the actual difficulty of the question. Therefore, we also analyzed activity for invalid vs. valid cues for easy and hard questions separately, and showed that during the confidence task, there was greater activity for validly cued hard questions than invalidly cued hard questions in the right angular gyrus, visual cortices, the right anterior cingulate, right putamen, and left occipital gyrus. This is in contrast to previous research on episodic recognition showing invalid cueing was associated with activity in the anterior cingulate, a region often associated with cognitive control (Botvinick, Cohen, & Carter, 2004), and the angular gyrus, which is thought to play a role in recollection during episodic retrieval tasks (Daselaar, Fleck, & Cabeza, 2006; O’Connor et al., 2010; Yonelinas et al., 2005). Therefore, we again see a dissociation between activity for invalid vs. valid cues in semantic and episodic recognition tasks.
For easy trials, there was greater activity for invalid than valid cues during the confidence task in the right caudate (BA48) and right cuneus (BA18). O’Connor et al. (2010) also showed that the cuneus and caudate were among the regions that showed greater activity during invalid than valid cueing during episodic recognition. There is some evidence that in addition to its involvement in visual processing, the cuneus may be also be engaged during inhibitory control (Chambers, Garavan, & Bellgrove, 2009), and there is converging evidence from neuroimaging, lesion, and non-human animal research that the caudate nucleus is involved in executive functioning, specifically the process of goal-directed behavior (De Simoni et al., 2018; Grahn, Parkinson, & Owen, 2008).

Taken together, it appears that the invalid cueing effect seen in episodic tasks was not replicated in our semantic task when questions were hard, but when questions were easy, there was some overlap between regions associated with invalid cueing for both tasks (O’Connor et al., 2010). Although these differences may reflect differences in invalid cueing effects for episodic as compared to semantic recognition, they could also stem from differences in recognition performance. Mean corrected recognition performance in the episodic memory task by O’Connor et al. (2010) was approximately 0.65, whereas in our task, overall proportion correct was only 0.45. Furthermore, performance differed significantly by question difficulty, with mean performance for easy questions being 0.64 and hard questions being 0.31. Thus, our finding that invalid cueing only for easy questions elicited similar activity as invalid cueing in the task by O’Connor et al. (2010), may reflect comparable recognition performance in these conditions, and suggests the effects of cue validity on brain activity varies with question difficulty.

MTL activity varies with individual differences in cue dependency

The prior analyses focus on the presence of the cue and its relationship to the question, but it is also important to consider the extent to which individuals used the cue to make their confidence judgments. Behavioral measures suggested that participants may have had higher confidence for invalidly cued hard questions answered correctly, as compared to uncued hard questions answered correctly. In the current experiment, an invalidly cued hard hit is a condition in which the subject is faced with a hard question, is cued to believe that it is easy, and successfully retrieves the correct response. Therefore, as opposed to confidence that stems from an internal indicating successful retrieval, which
would be equal for cued and uncued conditions, the inflation in confidence in this condition appears to be driven by the external cue signifying that other subjects also tend to answer the given question correctly. To examine how brain activity varies based on individual cue usage, we assessed the degree to which individual subjects incorporated the invalid cue into their confidence judgments by calculating the difference in mean confidence for invalidly cued hard hits – uncued hard hits, and used this individual cue dependency score as a covariate in the contrast of Confidence > Recognition. Activity associated with the process of making a confidence judgments significantly covaried with cue dependence in the bilateral medial temporal lobes, with peak activation in the amygdala extending to the anterior hippocampus. In other words, the more subjects’ RCJs were inflated by the invalid cue during hard hits, the more activity there was in the MTL during confidence judgments.

Activity in the MTL for correctly answered general knowledge questions is somewhat atypical, given that that although there is a significant body of literature implicating the MTL in episodic memory (Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012; Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996; Schacter & Wagner, 1999), its role is less frequently associated with semantic recognition. However, a few neuroimaging studies have implicated the MTL in semantic retrieval (Ryan, Cox, Hayes, & Nadel, 2008; Ryan, Lin, Ketcham, & Nadel, 2010; Verfaellie, Bousquet, & Keane, 2014), and semantic dementia has been associated with atrophy in MTL regions including the anterior fusiform and amygdala (Chan et al., 2001). In addition to potentially playing a role in semantic retrieval, research has suggested the MTL may also serve as a site of integration between multiple sources of information (van Kesteren, Ruiter, Fernández, & Henson, 2012). One model of MTL function suggests substantial MTL-mediated learning takes place during cases of prediction errors, due to the region’s role in integrating old and new experiences and knowledge (Shohamy & Wagner, 2008). Furthermore, there is evidence that the MTL is specifically engaged during encoding of information that is incongruent with existing schemas, such as encountering an unfamiliar item in a familiar context (van Kesteren et al., 2012). In the current study, an Invalidly Cued Hard Hit trial may require subjects to integrate internal cues that arise from the experience of accurate retrieval with new information in the form of an external cue about question difficulty, in order to accurately assess confidence. Thus we speculate that MTL activity associated with dependence on an invalid cue during confidence assessment for a hard question answered correctly could reflect integration
of multiple sources of information. O’Connor et al. (2010) found that invalid cueing during episodic retrieval was associated with activity in the anterior medial temporal lobe. This activity included a cluster in the right hippocampus which, although it was slightly more dorsal and posterior (21, -15, -12) than the MTL clusters we found, suggests regions typically associated with retrieval success may be influenced by subjects’ expectations based on external cues (O’Connor et al., 2010).

Limitations

An important potential limitation of the current study is the use of cues that were valid 50% of the time, raising the question of whether or not participants incorporated the cues into their confidence judgments. Previous research has used cues that are either largely reliable (e.g., 75% valid), random (50%), or largely unreliable (e.g., 25% valid) in order to determine how individuals incorporate external cues into recognition confidence judgments based on their utility (Jaeger, Cox, et al., 2012; Jaeger, Lauris, et al., 2012; O’Connor et al., 2010). In one such experiment, Jaeger, Lauris, et al. (2012) found that subjects’ responses conformed to both reliable (75% valid) and unreliable/random (50% valid) sources. Furthermore, recognition performance was equal for unreliably cued conditions and uncued conditions, which the authors suggest is due to subjects being more likely to conform to unreliable cues in instances of low confidence. In other words, in cases when subjects have low confidence because they are unable to rely on their own memory, in which case their performance might be at chance (50%), there is no risk in relying on a random external source that also has chance accuracy, because it will not impair their performance relative to guessing (Jaeger, Lauris, et al., 2012). Unlike the experiment conducted by Jaeger, Lauris, et al. (2012), which focused on the conformity of recognition responses to external cues, our experiment specifically tested the effects of external cues on confidence ratings as well; however, because our cues were valid 50% of the time, we were unable to test whether the effects of cues on confidence, and associated brain activity, may have varied with the reliability of the cue. However, Jaeger, Lauris, et al. (2012) found that subjects conformed to cues when they were random (50% valid) and even largely unreliable (25% valid), and similar research on collaborative recall has shown that subjects are just as likely to conform to the responses of confederates who provided 100% incorrect details as those who were only inaccurate 33% of the time (Numbers, Meade, & Perga, 2014). Therefore,
it appears that individuals often conform to external cues regardless of their reliability, supporting our finding that subjects were influenced by cue validity despite our use of cues that were 50% reliable.

Conclusions

Brain regions engaged when making confidence judgments during semantic recognition were similar to those associated with confidence assessment in episodic tasks, including the medial prefrontal, medial temporal, and medial and lateral parietal cortices (Chua et al., 2006, 2009b). Moreover, and delving deeper into the neurocognitive bases of retrospective confidence judgments, the presence of external cues about question difficulty modulated activity in regions associated with memory monitoring, particularly lateral and medial parietal and inferior temporal regions that may be involved in response uncertainty. Furthermore, there were effects related to the validity of these external cues that varied based on task difficulty, with greater activity for valid than invalid trials in the lateral parietal cortex, with peak activation in the angular gyrus, for hard questions, and greater activity for invalid than valid trials in the caudate and cuneus during easy questions. Finally, we found that the more individuals relied on an invalid cue during RCJs for correctly answered hard questions, the greater the activity in bilateral medial temporal lobes, which may be consistent with theories of the MTL as a point of integration between multiple sources of information (van Kesteren et al., 2012). The current study makes a novel contribution to our understanding of the brain regions involved in metamemory monitoring and semantic recognition by providing evidence that RCJ- and recognition-related activity differs when subjects make confidence judgments based on external information as compared to internal cues arising from the subjective experience of retrieval. Furthermore, we identified brain regions that respond to the presence and validity of external information on which individuals base confidence in semantic memory, which has been previously untested. Given past evidence that individuals likely rely more on external information when memory signals are weak, characterizing brain regions that respond to external cues may inform future hypotheses about how populations with memory deficits, particularly those associated with abnormal functioning of parietal and temporal regions, use external information to guide their metamemory monitoring.
Chapter 4: General Discussion

To our knowledge, these are the first studies to directly test how the roles of brain regions involved in metamemory monitoring vary with the information on which confidence judgments are made, both prospectively (i.e., with JOLs) and retrospectively. In both experiments, we tested the roles of brain regions previously shown to contribute to the process of making a subjective confidence judgment, the level of confidence expressed, and/or the accuracy of the judgment in relation to objective memory performance. Taken together, our results support the hypothesis that the neural mechanisms underlying metamemory monitoring during both encoding and retrieval vary with cognitive bases of the judgments. Furthermore, our results suggest some overlap between brain regions supporting metamemory monitoring and cognitive control processes, particularly when faced with multiple and sometimes conflicting sources of information on which to base confidence judgments.

There are many parallels between the predominant model of metamemory, in which subjects use the products of metamemory monitoring (the “meta level”) to exert control over memory processes (the “object level”) (Nelson & Narens, 1990), and leading models of cognitive control, in which an executive system monitors and regulates lower-level information stemming from environmental cues (Fernandez-Duque, Baird, & Posner, 2000; Shimamura, 2000). Past research has not only suggested that metacognitive regulation and executive control rely on many similar cognitive processes, but also identified brain regions that are common to both systems including anterior cingulate, aPFC, DLPFC, and basal ganglia (Fernandez-Duque et al., 2000). In Experiment 1, we used high definition transcranial direct current stimulation (HD-tDCS) to show that the aPFC, a region implicated in higher-order control of cognitive processes (Badre, 2008) has a causal role in JOL accuracy, but this role varies based on cues related to fluency of encoding; stimulation to this region increased JOL accuracy for low-frequency and inverted words, which are fluently encoded, but decreased accuracy for high-frequency words, which are fluently encoded. We also showed that the DLPFC plays a causal role in JOL accuracy, with stimulation decreasing accuracy specifically for high-frequency words, and in encoding success, with stimulation increasing subsequent false alarms.
In Experiment 2, we used fMRI to investigate how the roles of the prefrontal, parietal, and temporal cortices may change with the presence and validity of external cues about question difficulty in a semantic recognition task, and found that the presence of a cue altered activity in the parietal cortex, while the validity of cues influenced activity in the basal ganglia, which have previously been associated with error detection and goal-directed learning (Grahn et al., 2008; Holroyd & Coles, 2002). Taken together, the results of these experiments support the hypothesis that the brain regions involved in metamemory monitoring vary with the cognitive bases of the judgments, and make an important contribution to our understanding of metacognition in the brain by providing evidence for overlap between the neurocognitive mechanisms underlying metamemory and executive control.

The roles of the aPFC and DLPFC in JOLs and encoding

Experiment 1 showed that subjects tend to base their JOLs on cues that are not predictive of memory success, namely, fluency of encoding. JOL ratings were higher for high- as compared to low-frequency words, and upright as compared to inverted words, which is consistent with a significant body of previous work on the cognitive bases of JOLs. Past research has demonstrated that memory confidence increases with fluency of encoding (Besken & Mulligan, 2013; Koriat, 2008a; Miele et al., 2011), and that low-frequency words present a condition of disfluency due to lack of familiarity (Balota et al., 2002; Jia et al., 2016), while inverted words are perceptually disfluent, resulting in more effortful processing, resulting in lower confidence (Johnston et al., 1985; Rhodes & Castel, 2008). Past work has suggested disfluency is a ‘desirable difficulty’, in that conditions that challenge the learner during encoding lead to enhanced memory (Bjork, 1994), but subjects tend to discount this benefit when making JOLs (Koriat, 1997; Yue et al., 2012). Indeed, we found that despite lower JOLs for disfluent conditions, subjects had better recognition memory for both low-frequency and inverted words. This is consistent with work showing that memory performance is better under a variety of conditions of disfluency, such as enhanced memory for words that are masked so they are perceptually occluded as compared to clearly presented (Hirshman & Mulligan, 1991), when subjects assess semantic characteristics of verbal stimuli as compared to phonemic or physical characteristics (Craik & Tulving, 1975), and when words are presented once as compared to several times (Tulving & Kroll, 1995). Moreover, past research has
demonstrated that memory benefits from novelty of verbal stimuli (Kishiyama & Yonelinas, 2003; Kishiyama et al., 2009), and this may be due to deeper processing or attentional orienting to less familiar stimuli during encoding (Corbetta & Shulman, 2002; Kishiyama et al., 2009). Therefore, the effects of disfluency on memory and metamemory performance may also be interpreted as an effect of stimulus novelty, in that low-frequency words are novel due to unfamiliarity with the word itself, while inverted words are made novel by the format in which they are presented. Taken together, the behavioral results of Experiment 1 support a commonly demonstrated dissociation between the effect of fluency on memory and metamemory, wherein JOLs are higher with greater encoding fluency, but subsequent recognition performance benefits from encoding disfluency.

Although the cognitive bases of JOLs are relatively well established, neuroscience research has lagged behind, and it remains unclear how the brain supports JOL processes, particularly when JOLs are based on varying information. Experiment 1 showed that the DLPFC plays a causal role in encoding and JOL accuracy, the aPFC contributes to JOL accuracy, and the roles of both the DLPFC and aPFC on JOL accuracy vary based on encoding fluency. HD-tDCS over the DLPFC impaired encoding, as evidence by increased false alarms for all cue types. This is consistent with previous work demonstrating a role for the DLPFC in encoding that predicts subsequent false alarms (Demeter et al., 2016; Slotnick & Schacter, 2004). Increased activity in the lateral PFC during encoding predicts subsequent false alarms to familiar but unstudied lures (Slotnick & Schacter, 2004), and excitatory stimulation to the DLPFC during encoding lowers the proportion of false alarms to lure items at test (Demeter et al., 2016), suggesting a role of the DLPFC in the encoding of specific information that allows for discrimination between old and new events at test. As described above, subjects typically show a memory advantage for novel items at encoding, and this is demonstrated by fewer false alarms for inverted and low-frequency words following sham and aPFC stimulation. Although there was no significant interaction between stimulation and cue type on false alarms, pairwise comparisons did show that there was no significant difference between false alarms for high- and low-frequency words following DLPFC stimulation, suggesting stimulation eliminated the typical novelty advantage in recognition. One previous fMRI study showed that low-frequency words were associated with greater activity in the left prefrontal cortex during encoding, including the left inferior frontal gyrus and anterior cingulate, and similar regions showed greater activity for subsequent hits as
compared to misses on a recognition task. This implies the recognition benefit for low-frequency words may result from stronger encoding due to the recruitment of prefrontal regions that interact with medial temporal regions in support of successful memory formation (Chee, Westphal, Goh, Graham, & Song, 2003). Therefore, our finding that DLPFC stimulation increased false alarms is consistent with research implicating this region in processes at encoding that predict subsequent false alarms and, furthermore, this may be due to a disruption of novelty-based encoding for low-frequency words, as there was no difference in false alarms for low- and high-frequency words after DLPFC stimulation.

HD-tDCS to the aPFC selectively improved JOL accuracy for disfluent conditions, indicating the causal role of this region in JOL accuracy differs by cue type. In a TMS study of JOLs and associative recognition, Ryals et al. (2016) showed that theta-burst stimulation to the aPFC improved JOL accuracy relative to DLPFC and sham stimulation, and this was driven by a decrease in JOLs for subsequent misses, but no change in JOLs for subsequent hits. In other words, stimulation made subjects’ JOLs more sensitive to failures of encoding but did not alter JOLs for successfully encoded items. Conversely, in our study, JOL accuracy was increased specifically for inverted and low-frequency words, which were better remembered, and both DLPFC and aPFC stimulation impaired JOL accuracy for high-frequency words relative to sham. Therefore, it appears the role of the aPFC in JOL accuracy differs based on encoding success.

One theory of the roles of PFC subregions in cognitive control processes proposes that functions are arranged along the rostro-caudal axis of the PFC in a hierarchy of increasing complexity, wherein the aPFC guides concrete, lower-level signals from posterior regions of the PFC to coordinate inputs and internal states in the service of more abstract goals (Badre, 2008; Koechlin et al., 1999, 2003; Miller & Cohen, 2001). If anterior regions of the PFC integrate all lower-level information from posterior regions, it would follow that stimulation to this region would increase dependence on stimulus-related information, such as word frequency and orientation, regardless of their predictive value. In other words, for conditions in which subjects base JOLs on cues that are predictive of memory success, JOL accuracy would be increased by aPFC stimulation, whereas for conditions in which JOLs are inaccurate because they are based on nondiagnostic cues, stimulation would impair JOL accuracy due to increased reliance on these cues. However, we found that subjects had equally poor accuracy for all conditions under the
sham stimulation, suggesting subjects based JOL on nondiagnostic cues in all conditions (giving higher JOLs to fluent conditions and lower JOLs to disfluent conditions). Therefore, in the current study, if the aPFC integrated all cue information regardless of diagnosticity, stimulation to this region would make JOL accuracy worse for all conditions, because subjects base their JOLs on the nondiagnostic cue of fluency in all conditions. Instead, our results suggest that stimulation to the aPFC biases individuals toward using disfluency as a predictive cue of memory success, but also made JOLs less sensitive to the fact that fluency results in poorer memory. The finding that enhancing activity in the aPFC biased subjects toward using predictive cues is consistent with hierarchical models of PFC organization that suggest that when bottom-up input from posterior regions activates multiple possible responses, the aPFC acts in a top-down manner to bias responses in the interest of order task-relevant goals (Badre, 2008).

However, we also found that aPFC stimulation made subjects less sensitive to predictive cues for high-frequency words, suggesting that in this case, rather than biasing individuals toward giving lower JOLs to fluently encoded words, they actually became more dependent on the nondiagnostic cue of high fluency. Furthermore, stimulation to the DLPFC also reduced JOL accuracy for high-frequency words. In the context of the hierarchical model of prefrontal organization, this could reflect an increase in the salience of the non-diagnostic cue: if posterior regions of the PFC track stimulus-level externally-generated information (Badre, 2008; Christoff & Gabrieli, 2000), enhancing DLPFC activity could have increased the salience of high-frequency words and, consequently, made subjects more likely to base JOLs on the non-diagnostic cue of encoding fluency, reducing their accuracy. However, further research is needed to understand whether the DLPFC and aPFC interact in a hierarchical manner with regard to metamemory processes, and how manipulation of activity in each of these regions may alter cue usage to influence JOL accuracy.

---

**Brain activity associated with RCJs varies with external sources of information**

In Experiment 2, we showed that activity during retrospective confidence judgments was greater in medial prefrontal, and medial and lateral parietal cortices, as compared to the recognition task. Activity in these regions is consistent with past research examining the neural bases of RCJs as compared to recognition in episodic memory tasks (Chua et al., 2006, 2009b), as well as other forms of metamemory
monitoring, such as feeling-of-knowing judgments (Chua et al., 2009b), suggesting some shared neural mechanisms underlying metamemory processes across tasks. The regions in which we saw greater activity during confidence assessment are consistent with areas implicated in the “default mode network”, a group of brain regions that are coactivated during rest, and deactivated during tasks in which attention is directed to external stimuli (Gusnard et al., 2001; Raichle et al., 2001; Raichle & Snyder, 2007). Default mode network activity has also been broadly implicated in a variety of self-referential processes, including mind wandering (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009), mental simulation (Buckner & Carroll, 2007), and mindful awareness of one’s thoughts (Berkovich-Ohana, Glicksohn, & Goldstein, 2012; Tops, Boksem, Quirin, IJzerman, & Koole, 2014) suggesting an involvement of these regions in internally directed cognition. Therefore, activity in these regions during confidence judgments is consistent with evidence that the default mode network is engaged during self-referential processing.

There is some evidence that the default mode network may consist of multiple subsystems, including a medial temporal lobe system related to memory processing, and a medial prefrontal system that uses information from the MTL network in support of self-relevant cognition (Buckner, Andrews-Hanna, & Schacter, 2008). We found that the process of making an RCJ recruited brain regions associated with the prefrontal subsystem, supporting the hypothesis that the process of assessing one’s confidence engages regions that support self-referential processing.

A primary aim of Experiment 2 was to identify neural activity that varies with the presence of external cues during RCJs, and we found greater activity during uncued trials in the medial parietal and inferior temporal lobes, as compared to cued trials. The left medial parietal lobe has been associated with self-related processing, including judgments based on subjective preferences (Zysset et al., 2002) and the speed with which subjects retrieve self-referential episodic information (Lou et al., 2004). Therefore, we speculate that increased activity in these regions may reflect subjects relying more heavily on internal cues to make RCJ when faced with an uncued trial, given that these trials lack an external cue on which subjects can base confidence ratings. Past work has suggested subjects are more likely to rely on external sources of information when their internal memory signals are weak (Bradfield et al., 2002; Jaeger, Lauris, et al., 2012; Koriat et al., 2008; Wells & Bradfield, 1999). Although the finding that subjects rely on external memory cues when internal cues are weak does not directly inform the opposite
comparison (i.e. how subjects make recognition judgments in the absence of external cues), it does suggest subjects use a combination of information- and experience- based cues when assessing their confidence, and that in the absence of one type of cue, they will rely more heavily on other available sources of information. Therefore, it is plausible that medial parietal activity associated with uncued trials reflects greater reliance on internal experience-based memory signals during RCJs.

Another aim of Experiment 2 was to test how brain activity varies with the validity of external cues during RCJs in a semantic recognition task, and how this relates to past research on RCJs in episodic recognition. Here, we found that activity associated with cue validity during our semantic recognition task differed from cue validity effects shown in episodic recognition (O’Connor et al., 2010). No regions showed greater activity for invalid as compared to valid cues during RCJs or recognition, a finding which is inconsistent with the results of O’Connor et al. (2010), wherein invalid cueing in an old/new recognition task was associated with medial PFC and inferior parietal activity. This suggests the effects of cue validity on prefrontal and parietal activity vary based on memory modality. Other metamemory monitoring literature found that different brain regions are engaged for feelings of knowing based on episodic vs. semantic recognition. Maril, Simons, Mitchell, Schwartz, & Schacter (2003) found that as compared to ‘don’t know’ responses, feelings of knowing were associated with activity in the middle frontal gyrus during episodic recognition, whereas in a different study, Maril, Simons, Weaver, & Schacter (2005) found FOKs were correlated with greater activity in lateral and posterior medial parietal and superior prefrontal cortices for semantic recognition. Furthermore, a direct comparison of FOK-related activity during episodic and semantic retrieval within the same experiment showed greater activity during FOKs for semantic recognition in the right inferior frontal cortex, and greater activity for episodic FOKs in the middle temporal gyrus and posterior cingulate (Reggev et al., 2011). These results suggest a distinction between metamemory judgments based on episodic vs. semantic retrieval, and support our finding that the neural mechanisms underlying RCJs based on invalid cues during semantic recognition differ from those associated with invalid cueing during episodic recognition. Metamemory research has not typically addressed how confidence-related brain activity varies based on the memory modality and/or external sources of information (Reggev et al., 2011). Thus, our results represent a significant contribution to our
understanding of how subjects incorporate external cues into recognition and confidence judgments, and how this may differ for semantic as compared to episodic memory.

Because we investigated the neural bases of RCJs based on semantic recognition, we were able to use an objective marker of item difficulty, as opposed to episodic tasks, in which confidence is typically used as a marker of difficulty. We defined Easy and Hard questions by a median split in performance for general knowledge questions that had previously been answered by a large sample of college students, which allowed us to analyze how cue validity may have influence brain activity differently for questions that were objectively difficult vs. easy. Despite discrepant findings between our study and those of O’Connor et al. (2010) for the overall Invalid > Valid contrast, we found that the comparison of Invalid > Valid cues for Easy questions only showed greater activity for invalid cues in the right caudate and cuneus. Here, we show commonalities between our study and previous research showing these regions respond to invalidly cued trials in episodic memory (O’Connor et al., 2010). The caudate nucleus has been implicated in the processing prediction errors, i.e. how much an event differs from the expected outcome, in a broad range of non-human animal studies investigating reward-based learning (Grahn, Parkinson, & Owen, 2009; Haruno & Kawato, 2006; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004). In humans, there is some evidence that the caudate plays a similar role in more complex cognitive functions, guiding goal-directed behavior by monitoring how closely events match expectations. One fMRI study showed that when subjects were trained to produce a series of body movements in response to auditory cues, and then watched videos of dancers producing the same movements, the caudate nucleus showed greater activity when the dancers did not produce the expected movement for the given cue, suggesting this region tracks violations of expectations (Schiffer & Schubotz, 2011). Furthermore, it has been proposed that the error-related negativity (ERN), an even-related potential produced when subjects make errors in cognitive tasks, may reflect a prediction error signal stemming from the basal ganglia, including the caudate, which is then conveyed to executive control regions through functional connections with frontal regions (Holroyd & Coles, 2002). Therefore, our results, and those of O’Connor et al. (2010), showing greater caudate activity for invalid cueing, can be interpreted in the context of studies showing a role for the basal ganglia in tracking breaches of expectation, which in this case may occur when subjects are told a question is hard when in fact it is easy. The fact that
greater activity for invalid as compared to valid cues was only seen for easy questions in Experiment 2 may be due to a stronger conflict between internal and external memory signals. In the case of an easy question, subjects are likely to know the correct answer, which may elicit a strong internal experience-based memory signal. Therefore, when presented with an invalid external cue, there may be greater conflict between internal and external signals when assessing confidence. In contrast, for a hard question, subjects may not have a strong internal memory signal because they are less likely to know the correct answer, which consequently produces less conflict between internal and external cues.

Finally, we found that activity in bilateral medial temporal lobes was correlated with cue dependency for invalidly cued hard questions that were answered correctly. The MTL has typically been associated with episodic memory processes (Eichenbaum et al., 2012; Henson, 2005; Ranganath, 2010; Schacter & Wagner, 1999), but there is some evidence that it is also engaged during semantic recognition tasks (Chan et al., 2001; Ryan et al., 2008, 2010). Given that these studies implicate MTL involvement in retrieval of semantic memory, we might expect that MTL activity would be associated with the process of making RCJs based on internal memory signals as compared to external sources of information in our semantic recognition task. However, the more confidence ratings were inflated by the presence of an invalid external cue, the more activity we saw during RCJs in bilateral MTL. Therefore, we speculate that MTL activity here may reflect integration of multiple sources of information, e.g., external and internal cues about memorability. The MTL, and the hippocampus in particular, have been implicated in fMRI studies of retrieval-mediated learning, wherein memories of previous events are reactivated during the encoding of new related information (Shohamy & Wagner, 2008; Zeithamova, Dominick, & Preston, 2012), and one MEG study showed that theta band activity in the hippocampus during encoding predicted subsequent memory integration (Backus, Schoffelen, Szabényi, Hanslmayr, & Doeller, 2016). There is also evidence that hippocampal activity increases when subjects have to integrate serially presented information into a cohesive mental representation, such as during mental scene construction, in which subjects integrate discrete phrases describing elements of a scene into a unified representation of the scene (Summerfield, Hassabis, & Maguire, 2010). These studies suggest the hippocampus plays a role in linking memories for events that occurred at different times, but share overlapping content (Backus et al., 2016; Shohamy & Wagner, 2008; Zeithamova et al., 2012). During the
process of making an RCJ, subjects must integrate information related to internal cues, such as the strength of a memory signal during retrieval, with external cues about question difficulty, in order to decide the appropriate level of confidence to express. Although the above studies implicate the MTL specifically in the activation of previously encoded memory content during encoding of new information, the broader role of this region in integration of multiple sources of information may account for the increased activity we found in the MTL with greater dependency on invalid cues during hard hits. MTL activity may reflect integration between internal cues reflecting successful retrieval of the correct answer to a hard question, with the external cue indicating the question was easy. In fact, one model of hippocampal function suggests the hippocampus plays a specific role in integrating exteroceptive information, which stems from the external environment, with interoceptive information, such as emotional valence associated with episodic memories (Kassab & Alexandre, 2015). Therefore, activity in the hippocampus that increases with dependency on an invalid external cue during confidence assessment may reflect integration of internal and external memory signals. Furthermore, for invalidly cued hard questions, subjects inflated their confidence based on the presence of an external cue indicating the question was easy, but because the question was actually difficult, it may have required effortful retrieval, which would give rise to an internal experience-based signal of low confidence. There is some evidence that as part of its broader role in memory integration, the hippocampus is involved in comparing past and present experiences, and activity increases when what is expected to occur based on past experience conflicts with what occurs in the present reality (Kumaran & Maguire, 2006; Oehrn et al., 2015). Therefore, the increased MTL activity we saw with cue dependency during invalidly cued hard hits could reflect conflict between internal experience-based and external information-based cues. This finding extends previous work demonstrating a role for the MTL in integrating multiple sources of information, and provides new evidence suggesting this region may similarly contribute to metamemory monitoring by detecting and integrating signals from internal and external cues during RCJs, particularly when those cues conflict.
General Limitations

Experiment 1 used HD-tDCS to test the behavioral effects of directly manipulating brain activity in the PFC. Although there has been an increasing interest in the potential for tDCS to enhance functioning in a variety of clinical populations, and to test causal roles of cortical regions in cognitive processes (Colzato et al., 2017; Flöel, 2014; Nitsche et al., 2008; Zhao et al., 2017), there is also some controversy surrounding the ability of tDCS to alter performance in cognitive tasks (Horvath et al., 2015b; Jacobson et al., 2012). One meta-analysis of 42 studies using single-session tDCS during executive function, language, memory, and other cognitive tasks showed that tDCS had no effect on any of these functions. However, the results of this review have been called into question, particularly because the analyses combined studies that varied across dosage and duration of stimulation, and electrode montage (Antal, Keeser, Priori, Padberg, & Nitsche, 2015; Chhatbar & Feng, 2015). The question of how behavioral effects of tDCS relate to the dosage of current applied is important to our understanding of the mechanisms by which tDCS exerts its effects, and also pertain to our interpretation of the results of Experiment 1. Previous research has suggested that in addition to altering the conductance of ion channels to modify resting potential of neurons under the electrode (Stagg & Nitsche, 2011) tDCS may also exert its effects by modifying astrocytic calcium levels that contribute to synaptic plasticity (Monai et al., 2016). Small increases in calcium concentration have been shown to induce long-term depression, while large increases in calcium contribute to long-term potentiation (Stagg & Nitsche, 2011), highlighting the importance of current dosage in considering behavioral effects of tDCS.

Indeed, research has shown that the effects of tDCS on neural excitability vary with the duration and intensity of stimulation. One study found that when applying 1mA of current, cathodal tDCS inhibited excitability of neurons under the stimulation site, as evidenced by TMS-induced MEP amplitudes, whereas with 2mA of current, both cathodal and anodal stimulation enhanced excitability (Batsikadze, Moliadze, Paulus, Kuo, & Nitsche, 2013). Furthermore, past work has shown that doubling the duration of stimulation prolonged the excitatory after-effects of tDCS (Monte-Silva, Kuo, Liebetanz, Paulus, & Nitsche, 2010). In Experiment 1, we applied 2mA of stimulation, because this is the most commonly used strength of stimulation in cognitive tasks (Chua & Ahmed, 2016; Chua et al., 2017; Gaynor & Chua, 2016; Horvath et al., 2015a, 2015b; Hussey, Ward, Christianson, & Kramer, 2015; Nikolin, Loo, Bai, Dokos, &
Martin, 2015; Stagg & Nitsche, 2011), and previous research has shown this current intensity can alter metamemory performance (Chua & Ahmed, 2016; Chua et al., 2017). Furthermore, we limited stimulation to the duration of the encoding task (~10 minutes), because we were only testing online effects of tDCS, as opposed to after-effects which are more influenced by stimulation duration (Monte-Silva et al., 2010; Nitsche et al., 2005; Nitsche & Paulus, 2000). However, given the described studies demonstrating that neuronal excitability are sensitive to differences in the dosage and duration of current applied, future research should test how these factors influence heterogeneity in the behavioral effects of tDCS.

Because tDCS does not directly induce action potentials, but rather, alters resting membrane potential to make immediately subthreshold neurons more likely to fire (Bikson et al., 2004, 2012), the endogenous brain activity within the region being targeted by stimulation will also influence the dose-dependent response to tDCS. Indeed, a growing body of research has identified individual differences related to underlying neuronal activity, including mood states, education level (Berryhill & Jones, 2012), time of day (Gaynor & Chua, 2016; López-Alonso, Cheeran, Río-Rodríguez, & Fernández-del-Olmo, 2014), task difficulty (Chua et al., 2017; Jones & Berryhill, 2012), and baseline cognitive abilities (Jones & Berryhill, 2012; Tseng et al., 2012), may predict responses to stimulation. Additionally, variations in anatomy may also explain heterogeneity in the effects of tDCS on cognitive tasks. For example, studies have shown women had a significantly heightened response to tDCS of the visual cortex as compared to males (Chaieb, Antal, & Paulus, 2008), older subjects showed a delayed response to tDCS as compared to younger adults (Fujiyama et al., 2014), and differences in current density under the stimulating electrode due to anatomical differences influenced performance on a working memory task (Kim et al., 2014).

One hypothesis regarding variability in outcomes from tDCS suggests there is an optimal balance between cortical excitation and inhibition, and this balance varies between individuals. Therefore, an individual with intrinsically low excitation in a specific cortical region, may benefit from stimulation to that area; however, the same stimulation in a subject with high baseline excitability could lead to overexcitation which results in poorer performance (Krause & Cohen Kadosh, 2014; Krause, Márquez-Ruiz, & Kadosh, 2013). Therefore, there may be a curve that represents the relationship between neural
excitation and behavioral response, in which optimal excitability would fall in the center, with over- and under-excitability leading to suboptimal performance (Krause & Cohen Kadosh, 2014; Krause et al., 2013). Furthermore, if individual differences do indeed predict ‘responders’ and ‘nonresponders’ to tDCS, effects of tDCS at the group level can be obscured by analyses that group all subjects irrespective of individual differences, producing apparent null effects where in fact there may be a multimodal distribution of responses to stimulation (López-Alonso et al., 2014). In Experiment 1, we addressed this limitation by controlling for the effects of individual differences related to both anatomy (head size, age, gender), and factors shown to alter endogenous neural activity (self-reported mood state, alertness, difficulties concentrating). However, we did not test how duration and intensity of stimulation affected cognitive performance, and one possible future avenue of research is to manipulate current intensity and duration to test how individual differences in endogenous brain activity may interact with stimulation parameters to predict who does and does not benefit from tDCS.

Another potential limitation to Experiment 1 is the fact that during both aPFC and DLPFC stimulation, 17 of 24 subjects correctly guessed they were receiving active as compared to sham stimulation. Although sham stimulation involves a small amount of current administered at the start of stimulation to mimic the scalp sensations experienced during active stimulation, participant blinding during tDCS is known to be challenging due to differences in the duration and intensity of these sensations (Davis et al., 2013), and if subjects can indeed guess when they are receiving active vs. sham stimulation, this may impact behavioral results. However, we also found that 10 of the 24 subjects incorrectly believed they were receiving active stimulation during the sham condition, which could offset any differences seen in performance for active vs. sham conditions that might be due to the belief that subjects were receiving active current. Furthermore, there were no significant differences between aPFC and DLPFC conditions in terms of subjects’ guesses about whether they received active stimulation, which suggests the differences we saw in encoding success and JOL accuracy that were specific to each of these sites were unlikely to be influenced by participants’ beliefs about stimulation. Moreover, given that subjects are likely to base their guess about stimulation condition on the sensations they experienced during tDCS, our inclusion of covariates reflecting self-reported sensations during each session controlled for the possibility that their beliefs about stimulation condition impacted the observed results.
Nevertheless future studies could consider using a between-subjects design to limit the amount of experience each subject has with tDCS, or apply a topical anesthetic to reduce skin sensations associated with tDCS, to better blind participants to the stimulation condition.

In experiment 1, we stimulated the left DLPFC because previous studies have shown that tDCS to this region alters metamemory and memory performance in other tasks (Chua & Ahmed, 2016; Chua et al., 2017; Gaynor & Chua, 2016; Hussey et al., 2015), and the left lateral PFC has been widely implicated in episodic encoding (Epstein, Sekino, Yamaguchi, Kamiya, & Ueno, 2002; Fletcher, Shallice, & Dolan, 1998; Nyberg, Cabeza, et al., 1996; Spaniol et al., 2009). A significant amount of past research has also supported the idea that the left and right prefrontal cortices are differentially recruited during encoding and retrieval of episodic and semantic information, wherein the left PFC is specifically engaged during episodic encoding (Epstein et al., 2002; Habib, Nyberg, & Tulving, 2003; Nyberg, Cabeza, et al., 1996; Opitz, Mecklinger, & Friederici, 2000). Nevertheless, some studies have suggested that both right and left DLPFC are involved in episodic encoding; for instance right DLPFC activity has been associated with encoding of both verbal and nonverbal information (Opitz et al., 2000), and TMS to the right DLPFC during encoding has been shown to alter subsequent recall (Epstein et al., 2002) and recognition (Sandrini, Cappa, Rossi, Rossini, & Miniussi, 2003). Therefore, future research should address whether the role of the DLPFC in encoding and JOL accuracy as demonstrated in Experiment 1 is exclusive to the left hemisphere, or whether right DLPFC stimulation would produce comparable results.

One important limitation to the findings presented in Experiment 2 is the use of uncorrected p-values for the analyses comparing activity during Uncued vs. Cued and Invalid vs. Valid trials. Whereas we applied family-wise error corrections for multiple comparisons when analyzing the Confidence vs. Recognition contrasts, more specific analyses of activity by trial type were not corrected due to low power resulting from only 18 subjects in the sample, and the exploratory nature of the analyses. Although we used a conservative significance level (p<0.001) to limit false positives, the interpretation of these results should be taken in the context of increased probability of Type 1 errors due to not having been corrected for multiple comparisons. Nevertheless, the results of these analyses provide some preliminary evidence for regions that may be engaged based on the presence and validity of external cues, which may help
develop stronger hypotheses about regions of interest to be explored in future studies of brain activity associated with the incorporation of external cues during metamemory monitoring.

Another limitation to Experiment 2 relates to the use of general knowledge questions that were categorized as “Easy” and “Hard” based on a median split in recall performance from previously collected data. Some of these questions in the question database are considered “Hard” because they are answered incorrectly due to a lure in the form of a commonly endorsed incorrect answer, whereas others are difficult due to unfamiliarity with the domain or question content. We expected invalidly cued trials to present a potential conflict between internal signals arising from the subjective experience of question difficulty and the external cue about question difficulty. However, in the case of “Hard” questions with lures, although subjects tend to answer incorrectly, they may not experience a feeling of difficulty because they are confident in the incorrect answer and feel it is an easy question. Therefore, when these questions are presented with an invalid cue indicating they are “Easy”, there may be no conflict between the subjective experience of difficulty and the external cue. We included all trials in our analyses to maximize statistical power, but it may be fruitful to conduct future analyses that exclude Hard questions that were answered incorrectly due to subjects choosing the lure response, to more carefully control for the possibility that a portion of “Hard” questions were in fact subjectively experienced as “Easy” by most subjects.

In Experiment 2, we used fMRI to dissociate the brain regions that contribute to the process of making an RCJ from those engaged during the semantic recognition task. Although we found regions with greater activity for confidence than recognition, and vice versa, suggesting these processes are somewhat dissociable in the brain, it is worth noting that there may be confidence-related activity taking place during the recognition trial. The relationship between retrieval and metamemory monitoring is highlighted by a prominent theory of metacognition (Nelson & Narens, 1990), which suggests metamemory processes are divided between object and meta levels. Object-level information, such as the content of a memory, is monitored at the meta-level, and then controlled at the object level through changes in behavior. Practically speaking, this means that individuals are likely to control their recognition decisions based on feelings of confidence arising from the meta-representation of the memory. Indeed, studies have shown that memory search and termination strategies are influenced by
confidence: when subjects have a high feeling of knowing, (i.e. they feel they know the answer but can't currently retrieve it), they are more likely to continue searching their memory (Nhouyvanisvong & Reder, 1998). Therefore, it may be that in the case of RCJs, the process of confidence assessment cannot be entirely isolated from retrieval, because subjects may assess their confidence in retrieved information during the recognition trial, and use the results of that assessment to determine whether they should continue or terminate their memory search.

The issue of dissociating confidence from recognition processes may be partially eliminated through alterations to experimental design. One option is to use a block rather than event-related design: comparing brain activity during a block of recognition-only responses to a block of recognition and confidence judgments could help isolate the neural activity associated with confidence judgments. Alternatively, within an event-related design, we could include individual trials in which recognition is not followed by an RCJ, therefore providing a true recognition-only condition to which to compare activity associated with confidence assessment. Future work on the cognitive neuroscience of metamemory should consider these conceptual and methodological implications when testing for common and distinct neural components of memory and metamemory processes.

Another limitation of Experiment 2 is the use of a confidence scale with only 4 possible choices. Due to technological constraints of using a button-box to give responses in the MRI machine, we used a limited confidence scale ranging from 1-4, in which a response of “1” represented 0-25% confident, “2” represented 25-50% confident, “3” represented 50-75% confident and “4” indicated confidence of 75-100%. Because we were interested in testing the degree to which subjects incorporated cues into their confidence judgments, this study may have benefitted from a more fine-grained scale of confidence, for example, using a scale of 0-100% in 10% increments. Such a scale would be more sensitive to differences in confidence based on cue condition, which may have been obscured in Experiment 2 due to collapsing across large ranges of confidence. Although there are limitations to response capabilities in MRI, a joystick response rather than a button-box could allow subjects to give responses on a more fine-grained confidence scale. Despite this potential limitation, we did demonstrate significant differences in confidence between conditions, suggesting our 4-point confidence scale was sufficiently sensitive to evaluate how RCJs were affected by recognition performance, question difficulty, and external cues.
Broader implications

The results of this dissertation work inform our understanding of the cognitive and neural mechanisms that contribute to metamemory monitoring, and this has broader implications for other fields of study. Metamemory deficits have been demonstrated in populations with Alzheimer’s disease (Cosentino et al., 2016; Pappas et al., 1992; Souchay, 2007; Souchay et al., 2003), traumatic brain injury (Chiou et al., 2011; Kennedy, 2001, 2004), dementia (Cosentino & Stern, 2005; Souchay et al., 2003), schizophrenia (Moritz & Woodward, 2006; Souchay, Bacon, & Danion, 2006), depression (Kalska, Punamaki, Makinen-Pelli, & Saarinen, 1999), and various learning disabilities (Harris, Graham, & Freeman, 1988). In the learning domain in particular, accurate metamemory is crucial to the effective employment of study and testing strategies that facilitate optimal learning. JOL ratings have been correlated with strategic study decisions, wherein individuals choose to allot greater study time to items they previously gave lower JOLs (Son & Metcalfe, 2000), and individuals with better metamemory accuracy often show better memory performance and are more likely to correct their previous errors (Butterfield & Mangels, 2003). Therefore, an important question for future research is whether monitoring accuracy can be improved so that effective control strategies, such as appropriate study-time decisions, can be employed to benefit performance.

Studies have shown that metamemory deficits in populations with learning difficulties can be improved by enhancing metamemory monitoring and control abilities. For instance, Kennedy et al. (2003) showed that brain injury patients who demonstrate impairments in learning were able to make accurate metacognitive judgments by using delayed rather than immediate JOLs. Delayed JOLs, which are made after a determined period of time following the study of each item, tend to be significantly more accurate than immediate JOLs made during or directly after encoding (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Because delayed JOLs are likely made based on attempted retrieval from long-term memory (rather than short-term memory in the case of immediate JOLs), they are more predictive of long-term memory performance at later test (Nelson & Dunlosky, 1991). Kennedy et al. (2003) found that as compared to those who used immediate JOLs, TBI patients who used delayed JOLs were not only more likely to make accurate judgments, but then had improved recall following restudy of selected items because their choices about which items to study were based on more accurate JOLs (Kennedy, Carney,
& Peters, 2003). This suggests it is possible for individuals with learning impairments to overcome
deficits by learning to make metamemory monitoring and control decisions based on information that is
predictive of memory success, such as JOLs based on attempted retrieval.

Attention-deficit hyperactivity disorder (ADHD) has been widely associated with learning
difficulties in both children and adults, (Loe & Feldman, 2007; Mayes, Calhoun, & Crowell, 2000; Semrud-
clickeman et al., 1992; Woods, Lovejoy, & Ball, 2002) and studies show that individuals with ADHD are
also deficient in metamemory abilities (Antshel & Nastasi, 2008; Castel, Lee, Humphreys, & Moore, 2011;
Knouse, Anastopoulos, & Dunlosky, 2012). One study showed that adults with ADHD remembered fewer
words than controls, but were equally accurate in predicting memory performance using JOLs. However,
the ADHD group was less likely to employ study strategies, such as self-testing, that would improve their
memory (Knouse et al., 2012). Therefore, it is important to consider not only how individuals make
accurate metamemory monitoring judgments, but also the relationship between monitoring and control
strategies. Nevertheless, some studies have shown that training subjects in metamemory monitoring can
enhance self-regulated learning. For instance, JOLs are typically inflated for backward-associated word
pairs, in which the likelihood of the first word eliciting the second during cued recall (e.g., rain – umbrella)
is much lower than for a forward associated-pair (e.g., umbrella – rain) (Koriat & Bjork, 2006). However,
when subjects practiced studying and recalling lists of word pairs several times, they became more
sensitive to mnemonic cues that actually predicted recall ability, such as retrieval fluency, and JOLs
became more accurate. Furthermore, when subjects were explicitly taught that JOLs for backward-
associated word pairs tend to be inflated because the presence of the cue and target together misleads
subjects to believe the target will be easily retrieved when the cue is presented alone, JOL accuracy
improved (Koriat & Bjork, 2006). This evidence supports the possibility that metamemory monitoring can
be improved by training learners to be more sensitive to mnemonic cues that are predictive vs.
nonpredictive of memory performance. Moreover, following both these metamemory monitoring training
procedures, subject exhibited better metamemory control, allocating a greater amount of study time to
backward than forward-associated word pairs (Koriat & Bjork, 2006). Thus, making subjects aware of the
predictive cues that lead to memory accuracy enhanced both JOL accuracy and subsequent self-guided
learning strategies. Further research should address how training subjects to use diagnostic cues during
confidence assessment can be applied to other types of metamemory monitoring, such as RCJs, as well as how brain-based interventions such as tDCS, which have been shown to enhance learning in a wide variety of tasks (Coffman, Clark, & Parasuraman, 2014; Flöel, 2014; Flöel et al., 2012; Meinzer et al., 2014), can potentially enhance the benefit of metamemory training in order to promote more accurate monitoring and control strategies.

Another important avenue of research relates to the prevalence of metamemory impairments in populations with neurodegenerative diseases, particularly those characterized by a primary deficit in memory abilities, such as AD. Lack of accurate awareness in one’s own memory failures has been widely demonstrated in AD, and can have serious implications for disease treatment, including a failure to seek help early on, and make decisions about medication management (Cosentino, Metcalfe, Cary, De Leon, & Karlawish, 2011). However, diagnosis of metamemory deficits often relies on patient- or caregiver-report, which are subject to bias, highlighting the necessity of objective systematic testing of metamemory to understand how metamemory impairments relate to the neural substrates of AD (Cosentino & Stern, 2005).

Experimental work using objective measures of metamemory have shown that AD patients are impaired in JOL accuracy as compared to healthy controls (McGlynn & Kaszniak, 1991; Moulin, Perfect, & Jones, 2000), and in FOK accuracy using an episodic sentence memory task (Duke, 2000) and a paired associate task (Souchay, Isingrini, & Gil, 2002). Using a semantic memory task Pappas et al. (1992) found that AD patients were impaired in FOK accuracy, but performed similarly to controls in RCJ accuracy, suggesting the metamemory deficits found in AD may depend on both the type of memory being monitoring, and the type of judgment elicited.

Deficits in AD patients’ insight into their cognitive impairments has been associated with dysfunction of the prefrontal cortices in both neuropsychological (Mangone et al., 1991) and neuroimaging studies (Starkstein et al., 1995; Vogel, Hasselbalch, Gade, Ziebell, & Waldemar, 2005). In Experiment 1, we used tDCS to show that the PFC plays a causal role in JOL accuracy in healthy adults, but to date no studies have tested whether stimulation of the PFC may enhance metamemory monitoring abilities in the AD population. However, a growing number of studies have shown that non-invasive brain stimulation may be a promising tool by which to enhance other cognitive processes in AD, including...
improving visual recognition memory using tDCS (Boggio et al., 2009), and improving object and action naming using TMS (Cotelli, Manenti, Cappa, Zanetti, & Miniussi, 2008). These results, combined with evidence that metamemory monitoring can be improved by training individuals to base confidence on cues that are diagnostic of memory success (Koriat & Bjork, 2006), suggest a possible future avenue of research testing whether combining metamemory training with stimulation to enhance activity in regions known to contribute to metamemory accuracy, such as the PFC, may enhance metamemory abilities in patients with AD.

Furthermore, the results of Experiment 2 suggest that in healthy adults, the process of making RCJs during semantic retrieval engages regions associated with the default-mode network, and this is consistent with previous work identifying similar regions as contributing to episodic RCJs (Chua et al., 2004, 2006, 2009b, 2009a). Several studies have demonstrated that AD patients have abnormal default-mode network activity as compared to healthy controls (Greicius, Srivastava, Reiss, & Menon, 2004; Liu et al., 2008; Wu et al., 2011); moreover, AD patients show decreased functional connectivity between prefrontal and parietal cortices, regions which were not only engaged during the process of making RCJs in Experiment 2, but also showed varying activity based on the presence of external cues about question difficulty. Given the possibility that people with weaker memory signals, such as those with AD, may be more reliant on external information when assessing their retrieval confidence (Bradfield et al., 2002; Jaeger, Lauris, et al., 2012), one important avenue of future research would be to test how dysfunction in these brain regions may contribute to metamemory inaccuracy in AD patients, and whether this relates to a possible over-reliance on external cues about memorability due to primary memory deficits.

Taken together, the results of these dissertation experiments contribute to our understanding of how the neural and cognitive bases of metamemory interact, and suggest the ways in which brain regions support JOL and RCJ processes varies depending on the information on which metamemory judgments are made. Establishing the neural correlates of metamemory processes, how they relate to cognitive mechanisms involved in metamemory, and whether interventions such as tDCS can enhance these functions, is crucial to the development of appropriate and effective treatments that have the potential to benefit a wide variety of both healthy and clinical populations.
Figures

Figure 1. Brodmann Areas (BA) comprising the lateral (A) and medial (B) prefrontal cortex. BA10, the most anterior region of the prefrontal cortex, and BA11, which includes the ventromedial prefrontal and orbitofrontal cortices, have been associated with JOL processes.


Figure 2. Modeled current densities based on HD-tDCS montages for aPFC (A) and DLPFC (B) using HD-Explore (Soterix Medical, New York, NY).
Figure 3. Subjects had higher hit rates for low-frequency than high-frequency words under all stimulation conditions (A). Subjects had higher false alarm rates following DLPFC stimulation relative to sham and aPFC stimulation (B), but this effect did not differ by word frequency. Error bars reflect standard errors of the means.

Figure 4. Hit rates for inverted words were significantly greater than hit rates for upright words, but there were no effects of stimulation on hits (A). DLPFC stimulation increased false alarms relative to sham and aPFC stimulation, but there was no effect of orientation on false alarms (B). Error bars reflect standard errors of the means.
Figure 5. Relative to sham and DLPFC stimulation, aPFC stimulation improved JOL accuracy for low-frequency words; aPFC and DLPFC stimulation impaired JOL accuracy for high-frequency words (A). Relative to sham and DLPFC stimulation, aPFC stimulation improved accuracy for inverted words, and there was a significant difference between inverted and upright words with aPFC stimulation (B). Error bars reflect standard errors of the means.
Figure 6. In the scanner, subjects answered and gave confidence ratings for 300 general knowledge questions, presented with cues about difficulty ("Easy", "Hard" or Uncued), which were valid 50% of the time based on actual question difficulty.
Figure 7. Significant variation in individual subjects' cue dependency scores for Invalidly Cued Hard Hits (mean RCJ for Invalidly Cued Hard Hits – mean RCJ for Uncued Hard Hits). Each dot represents an individual subject's cue dependency score. Upper and lower boundaries of the box indicate the 1st and 3rd quartiles of the data set, respectively, with the 2nd quartile (median) represented by the inner band. Whiskers indicate minimum and maximum values within the
Figure 8. The contrast of Confidence > Recognition showed greater activity in the medial prefrontal cortex (A, B) and medial parietal cortex (A). Colorbar represents t-values. Timecourse of signal change from cluster in medial PFC for confidence and recognition trials (C).
Figure 9. Comparisons of activity during Recognition > Confidence tasks showed greater activity in the medial prefrontal cortex (A, B), and left lateral prefrontal cortex (B). Colorbar represents t-values. Timecourse of signal change from cluster in left lateral PFC for recognition and confidence trials (C).
Figure 10. The contrast of Uncued > Cued trials during recognition showed greater activity for Uncued trials in the left lateral prefrontal cortex (A), and left lateral and medial parietal cortices (B). Colorbar represents t-values.
Figure 11. Differences in activity for Confidence > Recognition covaried with individual differences in Invalid Hard Hit cue dependency in bilateral MTL regions (A). Mean percent signal change in the left MTL (-12, -7, -17) increased with cue dependency (B). Mean percent signal change during Recognition did not appear to be related to cue dependency (C), but there was a positive relationship between activity during Confidence and cue dependency (D). A similar pattern of percent signal change was seen in the right MTL cluster. Datapoints represent mean percent signal change for all voxels within significant clusters; colorbar represents t-values.
Tables

Table 1. Clusters showing significant differences between the Recognition and Confidence tasks (p<0.001, uncorrected; 10 voxel extent threshold). Peak differences in activation are reported in MNI coordinates. Sub-peaks are reported when clusters span multiple Brodmann areas (BA) and hemispheres, and are further separated out if sub-peaks are located in different gross anatomical areas.

<table>
<thead>
<tr>
<th>Region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>No of voxels</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confidence &gt; Recognition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal cortex</td>
<td></td>
<td></td>
<td></td>
<td>187</td>
<td>5.63</td>
</tr>
<tr>
<td>BA10</td>
<td>0</td>
<td>50</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA11</td>
<td>6</td>
<td>41</td>
<td>-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA32</td>
<td>0</td>
<td>53</td>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial parietal cortices</td>
<td></td>
<td></td>
<td></td>
<td>135</td>
<td>5.23</td>
</tr>
<tr>
<td>BA 23</td>
<td>0</td>
<td>-22</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA24</td>
<td>3</td>
<td>-19</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA31</td>
<td>3</td>
<td>-31</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA31</td>
<td>-6</td>
<td>-25</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA23</td>
<td>-3</td>
<td>-31</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R lateral tempo-parietal cortex</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>5.05</td>
</tr>
<tr>
<td>R BA21</td>
<td>60</td>
<td>-28</td>
<td>-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R medial parietal cortex</td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>4.97</td>
</tr>
<tr>
<td>R BA23</td>
<td>12</td>
<td>-52</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recognition &gt; Confidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal cortex</td>
<td></td>
<td></td>
<td></td>
<td>51</td>
<td>5.08</td>
</tr>
<tr>
<td>L BA 8</td>
<td>-6</td>
<td>17</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA6</td>
<td>-3</td>
<td>14</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA6</td>
<td>6</td>
<td>14</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L lateral prefrontal cortex</td>
<td></td>
<td></td>
<td></td>
<td>62</td>
<td>5.45</td>
</tr>
<tr>
<td>L BA 44</td>
<td>-48</td>
<td>14</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA9</td>
<td>-45</td>
<td>26</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Clusters showing significant differences between Uncued and Cued trials during recognition task and confidence task (p<0.001, uncorrected; 10 voxel extent threshold). Peak differences in activation are reported in MNI coordinates. Sub-peaks are reported when clusters span multiple Brodmann areas (BA) and hemispheres, and are further separated out if sub-peaks are located in different gross anatomical areas.

<table>
<thead>
<tr>
<th>Region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>No of voxels</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L anterior prefrontal cortex (BA 10)</td>
<td>-24</td>
<td>65</td>
<td>10</td>
<td>25</td>
<td>3.76</td>
</tr>
<tr>
<td>R medial prefrontal cortex (BA 10)</td>
<td>9</td>
<td>59</td>
<td>7</td>
<td>11</td>
<td>3.36</td>
</tr>
<tr>
<td>L medial prefrontal cortex (BA 8)</td>
<td>-9</td>
<td>47</td>
<td>46</td>
<td>26</td>
<td>4.16</td>
</tr>
<tr>
<td>R medial prefrontal cortex (BA 8)</td>
<td>18</td>
<td>32</td>
<td>43</td>
<td>17</td>
<td>3.73</td>
</tr>
<tr>
<td>L lateral prefrontal cortex</td>
<td></td>
<td></td>
<td></td>
<td>129</td>
<td>4.04</td>
</tr>
<tr>
<td>L BA8</td>
<td>-33</td>
<td>14</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA6</td>
<td>-39</td>
<td>11</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L lateral prefrontal cortex (BA 8)</td>
<td>-42</td>
<td>14</td>
<td>31</td>
<td>11</td>
<td>3.65</td>
</tr>
<tr>
<td>L premotor cortex (BA 6)</td>
<td>-21</td>
<td>14</td>
<td>61</td>
<td>11</td>
<td>3.49</td>
</tr>
<tr>
<td>R motor cortex (BA 4)</td>
<td>51</td>
<td>-7</td>
<td>28</td>
<td>36</td>
<td>3.62</td>
</tr>
<tr>
<td>L motor cortex (BA 4)</td>
<td>-33</td>
<td>-7</td>
<td>34</td>
<td>39</td>
<td>3.75</td>
</tr>
<tr>
<td>R motor &amp; somatosensory cortices</td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>3.65</td>
</tr>
<tr>
<td>R BA4</td>
<td>18</td>
<td>-22</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA1</td>
<td>21</td>
<td>-34</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L middle temporal pole (BA 38)</td>
<td>-30</td>
<td>17</td>
<td>-29</td>
<td>10</td>
<td>3.90</td>
</tr>
<tr>
<td>L lateral temporal lobe (BA 21)</td>
<td>-51</td>
<td>-28</td>
<td>-5</td>
<td>19</td>
<td>3.51</td>
</tr>
<tr>
<td>Region</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>MAF</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>R lateral parietal lobe (BA 7)</td>
<td>39</td>
<td>-43</td>
<td>43</td>
<td>29</td>
<td>3.85</td>
</tr>
<tr>
<td>R inferior parietal cortex</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>3.73</td>
</tr>
<tr>
<td>R BA39</td>
<td>48</td>
<td>-61</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L superior parietal lobe (BA 7)</td>
<td>-18</td>
<td>-61</td>
<td>55</td>
<td>38</td>
<td>3.65</td>
</tr>
<tr>
<td>R lateral parietal lobe (BA 7)</td>
<td>30</td>
<td>-61</td>
<td>52</td>
<td>30</td>
<td>3.43</td>
</tr>
<tr>
<td>R medial parietal lobe</td>
<td></td>
<td></td>
<td></td>
<td>291</td>
<td>4.30</td>
</tr>
<tr>
<td>R BA7</td>
<td>3</td>
<td>-61</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R BA31</td>
<td>0</td>
<td>-49</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R visual association area¹ (BA 19)</td>
<td>0</td>
<td>-82</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R primary visual cortex (BA 17)</td>
<td>21</td>
<td>-67</td>
<td>10</td>
<td>27</td>
<td>3.73</td>
</tr>
<tr>
<td>R visual association area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>370</td>
</tr>
<tr>
<td>L BA19</td>
<td>-36</td>
<td>-70</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA17</td>
<td>-12</td>
<td>-82</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA18</td>
<td>0</td>
<td>-94</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA18</td>
<td>-21</td>
<td>-67</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L BA39</td>
<td>-30</td>
<td>-79</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L inferior parietal lobe (BA 37)</td>
<td>-30</td>
<td>-49</td>
<td>-14</td>
<td>23</td>
<td>3.56</td>
</tr>
<tr>
<td>L superior parietal lobe (BA 7)</td>
<td>-9</td>
<td>-58</td>
<td>10</td>
<td>25</td>
<td>3.93</td>
</tr>
<tr>
<td>L cerebellum</td>
<td>-6</td>
<td>-55</td>
<td>-26</td>
<td>43</td>
<td>4.28</td>
</tr>
</tbody>
</table>

¹ sub-peaks contiguous with 291 voxel cluster

Uncued > Cued during confidence

<table>
<thead>
<tr>
<th>Region</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R premotor cortex (BA 6)</td>
<td>27</td>
<td>2</td>
<td>52</td>
<td>14</td>
</tr>
<tr>
<td>L medial temporal lobe (BA36)</td>
<td>-18</td>
<td>-22</td>
<td>-8</td>
<td>12</td>
</tr>
<tr>
<td>L inferior temporal lobe (BA 37)</td>
<td>-30</td>
<td>-49</td>
<td>-14</td>
<td>23</td>
</tr>
<tr>
<td>L medial parietal lobe (BA 23)</td>
<td>-9</td>
<td>-58</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>L cerebellum</td>
<td>-6</td>
<td>-55</td>
<td>-26</td>
<td>43</td>
</tr>
</tbody>
</table>
Appendix A

Post tDCS questionnaire

Subject ID: ________________ Date: ________________

Notes on tDCS protocol:

<table>
<thead>
<tr>
<th>Severity</th>
<th>Relationship</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Absent</td>
<td>1- None</td>
<td></td>
</tr>
<tr>
<td>2- Mild</td>
<td>2- Remote</td>
<td></td>
</tr>
<tr>
<td>3- Moderate</td>
<td>3- Possible</td>
<td></td>
</tr>
<tr>
<td>4- Severe</td>
<td>4- Probably</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5- Definite</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did you experience any of the following symptoms/side effects?</th>
<th>Severity</th>
<th>Relationship</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headache</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck Pain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalp Pain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tingling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burning sensation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin redness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleepiness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trouble concentrating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Mood Changes (Indicate Direction)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (specify)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you think you received active or sham stimulation? ________________
References


https://doi.org/10.1111/ejn.12307


https://doi.org/10.1093/brain/awx309


https://doi.org/10.3758/s13415-016-0426-3


https://doi.org/10.1016/j.cognition.2010.09.012


https://doi.org/10.1016/j.neuropsychologia.2012.07.037


https://doi.org/10.1016/S0304-3940(01)02573-3


https://doi.org/10.1006/ccog.2000.0447


https://doi.org/10.1093/brain/121.7.1239


https://doi.org/10.1016/j.neurobiolaging.2011.05.007


https://doi.org/10.1080/17588928.2016.1213713

https://doi.org/10.1016/j.pneurobio.2008.09.004

https://doi.org/10.1016/j.bbr.2008.11.020

https://doi.org/10.1073/pnas.0308627101

https://doi.org/10.1073/pnas.071043098

https://doi.org/10.1016/S1364-6613(03)00110-4


Horvath, J. C., Forte, J. D., & Carter, O. (2015a). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP
amplitude modulation in healthy human subjects: A systematic review. 

*Neuropsychologia, 66*, 213–236. https://doi.org/10.1016/j.neuropsychologia.2014.11.021


https://doi.org/10.1016/j.brainres.2016.10.005


https://doi.org/10.1016/j.brainres.2009.11.047


https://doi.org/10.1037/a0028885


https://doi.org/10.1037/0278-7393.26.1.204


https://doi.org/10.1016/S0010-9452(08)70696-8


https://doi.org/10.1076/neur.9.6.482.29376

https://doi.org/10.1016/j.neuropsychologia.2009.02.028


https://doi.org/10.1177/1073858410386614

https://doi.org/10.1523/JNEUROSCI.2877-06.2006

https://doi.org/10.1001/archneur.1995.00540280105024


https://doi.org/10.3758/BF03212976

https://doi.org/10.1037/0022-0663.95.1.66

https://doi.org/10.3389/fpsyg.2014.00429


https://doi.org/10.3758/BF03202594


https://doi.org/10.1016/j.clinph.2015.11.012


