The good, the bad, and the ugly of top-down executive control

By
Taraz Geola Lee

A dissertation submitted in partial satisfaction of the
Requirements for the degree of
Doctor of Philosophy
in
Psychology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Mark D'Esposito, Chair
Professor Rich Ivry
Professor Robert T. Knight
Professor Michael Silver

Spring 2012
The good, the bad, and the ugly of top-down executive control

by

Taraz Geola Lee

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Mark D'Esposito, Chair

Theories of working memory and attention postulate that the prefrontal cortex (PFC) provides top-down signals to other brain regions in order to keep behaviorally relevant sensory information activated and to suppress competing task-irrelevant information. Although there are numerous studies that suggest the PFC is a source of top-down modulation of posterior brain regions, the vast majority of these studies offer only indirect evidence in support of this claim. Additionally, while this executive control is usually thought of as being beneficial to the task at hand, there is reason to believe that there are certain circumstances in which this control is deleterious to performance. Here we provide direct evidence using transcranial magnetic stimulation and neuroimaging that 1) the PFC is a source of top-down control of early visual regions and 2) interfering with PFC function can sometimes lead to paradoxical improvements in task performance.
Chapter 1. Introduction

1.1 Background

Even the most primitive and simple organisms can display routine and stereotyped behavior in response to incoming stimuli. This “bottom-up” processing is necessary and quite advantageous in order to act quickly and automatically in commonly faced situations. As one moves up the evolutionary ladder, however, animals have developed larger and larger brains and have acquired the ability to perform more complex and flexible behavior to deal with novel and dangerous situations in accordance with internal needs and goals. It is often necessary to override automatic tendencies or to coordinate lower-level processes for effective action. Additionally, the selection of the most appropriate course of action in a given situation often involves a countless number of control processes. One must continuously pay attention to the relevant sensory information from the environment, disregard irrelevant and distracting information, retrieve the necessary information from long-term memory, integrate and manipulate all this material, and then select the appropriate action based on this information. This “top-down” processing is often called “executive control” and is supported by a network of different brain regions that crucially includes the prefrontal cortex (PFC).

A vast array of studies have provided evidence that the PFC is important for top-down executive control. For over a century, neuropsychologists have observed that patients with damage to the frontal lobes have specific deficits in the planning and organization of behavior (Bianchi, 1895; Luria, 1973). Advances in neuropsychological assessments over the years has led to a departure from anecdotal evidence to the delineation of a wide variety of tasks from rule learning to task switching to long term memory in which prefrontal patients show concrete deficits (Milner, 1963; Petrides, 1982; Shimamura et al., 1991). Additionally, anatomical tracing studies in monkeys have consistently shown that the prefrontal cortex has reciprocal connections with virtually the entire brain (Petrides and Pandya 1984, 1999, and 2002). This has led many to further accept its role in executive control as it is a region that is ideally situated to provide top-down bias signals to other regions in service of task goals (Miller and Cohen 2001).

1.2 The Good

One set of processes that intimately relies on top-down executive control are those needed in service of working memory. Working memory has been described as the ability to guide behavior by maintaining and manipulating information that is no longer in the external environment (Baddeley, 1986). The capacity of working memory is limited (Cowan, 2005) and it is necessary to filter out competing task-irrelevant information in order to effectively encode task-
relevant sensory information for further maintenance (Vogel et al. 2005). Many studies have suggested that in support of working memory, the prefrontal cortex (PFC) provides top-down signals to the extrastriate cortex and other visual processing regions in order to selectively enhance the processing of relevant sensory information and to keep the neural representations of this information active even after it is no longer in the environment (Miller and D'Esposito 2005).

Working memory tasks often use a “delay” period over which information must be maintained before a response is given. In a typical delayed recognition task, a stimulus is briefly presented to the subject during an initial cueing period. After the stimulus is removed, information about this stimulus must be held in mind during a delay with no further stimulus presentations. At test, the subject either has to report the stimulus that was initially viewed or is asked to respond whether it matches a new probe stimulus. A number of studies have shown that the PFC displays persistent activity over such delays using single-unit recording in nonhuman primates (Fuster and Alexander, 1971; Miller et al. 1996) and using functional magnetic resonance imaging (fMRI) in humans (Curtis and D'Esposito, 2003). In extrastriate cortex, studies using electrophysiological recordings (Fuster and Jervey, 1982; Miyashita and Chang, 1988) and fMRI (Postle et al., 2003; Ranganath et al., 2004) have observed similarly elevated delay period activity. Specifically, there have been a large number of electrophysiology and neuroimaging studies that have provided evidence that neural activity is enhanced in those regions of extrastriate cortex that encode relevant stimuli (Fuster, 1990; Duncan et al., 1997; Kanwisher et al., 1998) and suppressed in visual regions that represent irrelevant stimuli (Kastner et al, 1998; Gazzaley et al., 2005). Evidence from tract-tracing studies in monkeys (Ungerleider et al., 1989; Webster et al, 1994) and human neuroimaging studies (Kastner and Ungerleider, 2000; Corbetta and Schulman, 2002; Gazzaley et al., 2007) suggest that the PFC is a source of this modulation in extrastriate activity.

Unfortunately, the mechanisms by which the PFC acts on other brain regions in support of top-down control in working memory remains elusive (Miller and D'Esposito, 2005). While many groups seek to better characterize these mechanisms using electrophysiological recordings and neuroimaging techniques, the vast majority offer only correlational evidence between PFC activity and downstream changes in neuronal activity and behavior. In recent years, several groups have begun combining multiple neuroscience techniques to directly detect causal relationships between the PFC and the rest of the brain. Chapter 2 of this dissertation provides an overview of the extant literature that has provided causal evidence for the role of the PFC in the top-down control of other brain regions. Each of the studies reviewed has combined electrophysiological/neuroimaging recordings with interventional techniques that either examines how PFC damage or directly stimulating prefrontal neurons affects processing in other brain regions. The discussion of these results will
focus on PFC control mechanisms that could only be revealed by the combination of multiple techniques.

Two techniques that are ideally suited to provide this causal evidence for the prefrontal control over early visual regions are transcranial magnetic stimulation (TMS) and fMRI. TMS is a technique used for noninvasive stimulation of the brain (for a review, see Hallett, 2007). Passing electric current through a magnetic coil generates a brief, high-intensity magnetic field that can cross the skull to excite or inhibit circumscribed areas of the brain below the coil. Because this technique can be used in healthy adult individuals, it provides several advantages over studying patients with brain lesions. Following injury, the cerebral cortex undergoes rapid reorganization to compensate for lost abilities. Conclusions about the function of damaged tissue drawn from patient populations may therefore be inaccurate as it can be difficult to ascertain behavioral changes due to lost tissue relative to reorganized healthy tissue. Moreover, it is often difficult to recruit sufficiently large patient populations with lesions localized to the same brain region. In contrast, TMS allows the selection of specific brain regions for stimulation a priori. Several groups have sought to combine TMS with electroencephalography and fMRI to not only examine the behavioral effects of stimulation, but also to determine changes in distal brain regions and in large-scale brain networks (Ruff et al. 2006; Taylor et al. 2007; Zanto et al. 2011; Higo et al. 2011; Feredoes et al. 2011)

The superior spatial resolution of fMRI relative to other imaging techniques, lends itself to the examination of whole-brain effects of focal brain stimulation during executive control tasks. In Chapter 3, I describe a study in which we use TMS to transiently disrupt the function of a region of the PFC just prior to the performance of a working memory in the fMRI scanner. By combining TMS with fMRI we reveal specific mechanism by which the PFC is causally able to maintain the selectivity of extrastriate cortex in service of working memory. Additionally, we provide evidence that disruption to the PFC in one hemisphere leads directly to compensatory activity from PFC in the opposite non-stimulated hemisphere to help maintain task performance.

1.3 The Bad and the Ugly

While it is often assumed that more executive control and monitoring of ongoing activity by the PFC leads to better performance on a variety of behavioral tasks, there have been several studies that suggest that certain types of executive control can often interfere with superior performance, paradoxically leading to impaired ability on certain tasks. In these reports, it is often shown that populations with impaired or overtaxed executive function can outperform healthy adults. While rare, there have been several studies that have suggested that patients with damage to select brain regions can sometimes outperform healthy adults in carefully controlled studies. In one such study, patients with lateral
prefrontal cortex damage are shown to outperform healthy adult individuals on a certain class of insight problem solving puzzles (Reverberi et al., 2005).

Young children comprise another population that lack full executive function capabilities. It has been shown by several groups that the PFC is the slowest region of the cerebral cortex to fully mature (Huttenlocher, 1979; Giedd, 2004). Despite this “deficit”, young children show a similar advantage in certain types of insight problem-solving like the study described above (German and Defeyter, 2000). Additionally, it is commonly observed that children seem to learn languages at a faster rate than adults (Lenneberg, 1967; Birdsong and Molis, 2001). Children under the age of 4 also tend to select more optimal strategies than older individuals in binary choice tasks (Derks and Paclisanu, 1967). It has been proposed that a potential explanation for these results in children is that they lack the cognitive control mechanisms which tend to interfere with the type of learning processes that lead to optimal outcomes in these scenarios (Thompson-Schill et al., 2009).

A more common approach to examining how executive control can interfere with superior performance has been to behaviorally manipulate attentional demands to examine how distraction impacts performance on a secondary task. It is well established that the addition of a secondary, extraneous task usually leads to performance decrements on a primary task (Pashler and Johnston, 1989). This finding is often explained in light of the capacity limitations of working memory acting as a “bottleneck” through which only a certain amount of effective processing can occur (Navon and Miller, 2002). Nevertheless, there have been several reports of paradoxical improvements of behavior on a primary task of interest in dual-task scenarios. Filoteo and others (2010) provided evidence that the concurrent performance of a sequential working memory task during study can improve specific types of category learning. Similarly, Beilock and others (2002) demonstrated that a secondary auditory discrimination task could improve the putting performance of expert golfers. There has also been evidence that performance on a target-detection task that would seem to require vigilant attention can be improved by a concurrent distracting task (Olivers and Nieuwenhuis, 2005). Finally, a recent pair of studies described conditions under which visual memory for abstract images can be improved if studied during dual task conditions (Voss et al., 2008; Voss and Paller 2009).

Moreover, some groups have provided evidence that under certain conditions, performance in single-task scenarios can suffer when the focus of attention is shifted to explicitly monitor function. Skilled typists produce more errors and type more slowly when asked to attend to the movements of their hands (Logan and Crump, 2009). Likewise, older individuals show impaired learning on sequential learning tasks when explicitly cued to discover the underlying pattern that is present in the task (Howard and Howard, 2001). Amongst subjects who are tested on the structure an artificial grammar after brief amounts of exposure, those who are given explicit instructions to search for the structure perform worse
than those subjects who are given more neutral instructions (Reber, 1989). These examples suggest that executive control can hamper superior functioning.

Unfortunately, there have been relatively few instances in which neuroscientific techniques are employed that can elucidate the neural mechanisms behind this type of interference by top-down control networks. One of the aims of this dissertation is to show that prefrontal control processes can sometimes impair task performance by providing direct evidence that disrupting the function of the PFC via TMS can lead to behavioral improvements. In Chapter 4, I target two different PFC sites with TMS prior to the performance of a visual memory task and show that recognition performance can be improved without awareness. Chapter 5 describes another study in which I provide evidence that disruptive PFC TMS can paradoxically ameliorate performance on the Stroop task, which has previously been shown to rely on executive control processes.
Chapter 2. On the necessity of combining neural perturbation with neuroimaging and electrophysiology in understanding working memory

2.1 Overview

As discussed in Chapter 1, the prefrontal cortex (PFC) is thought to be a source of top-down signals that are able to modulate activity in extrastriate cortex in the service of working memory. Unfortunately, the vast majority of studies suggesting that the PFC interacts with extrastriate cortex offer only indirect evidence in support of this claim. Often the neural activity in disparate brain regions, as measured by a variety of techniques, are examined and/or correlated in order to describe network activity. However, this work can does not provide the strongest evidence for an interaction between these regions and simply shows that these regions show common levels of activity across time. The causal role of the PFC and the precise nature of the top-down signals it sends to other cortical regions cannot be shown using these techniques alone.

In a recent review, Miller and D'Esposito (2005) highlighted the lack of direct evidence for PFC involvement in modulating unimodal association cortex activity and suggested that new multivariate statistical analyses of fMRI data might be able to improve our understanding of top-down neural interactions. Since then, several studies of top-down modulation have used such multivariate techniques including Granger causality mapping to infer directionality from their neuroimaging results (e.g., Bressler et al., 2008 and Gregoriou et al., 2009).

While this ever-increasing group of neuroimaging studies seeking to understand the role of the PFC has undoubtedly contributed to our understanding of how the brain supports working memory, imaging techniques alone are inherently correlative rather than causal. Although multivariate techniques such as those described above do provide some evidence of directionality, they rely on multiple mathematical and physiological assumptions and still provide indirect evidence of causality. Several groups have sought to combine neuroimaging and neuronal recording with techniques that look at disrupted and/or stimulated brain tissue as a complementary approach to understanding the network activity underlying working memory performance. These studies not only allow the examination of the causal impact of prefrontal regions on remote cortical regions, but they also reveal the more specific information about the timing, the context, and the mechanisms by which the PFC exerts its control. Here we discuss the emerging approach of combining perturbation and recording techniques and highlight its use in gaining a more nuanced understanding of top-down control processes in visual working memory and attention.
2.2 Top-down modulation of response magnitudes

Previous studies have shown that the magnitude of visual responses of neurons in early visual areas such as area V4 can be modulated by attention (Moran and Desimone 1985; Motter 1993). The frontal eye fields (FEF) are one of several frontal sites that have been implicated as a potential source of top-down modulation of these early visual regions (Corbetta et al., 1998). In one of the first causal tests of this assertion, Moore and Armstrong (2003) electrically stimulated sites within the FEF while recording activity from V4 neurons. The visual response in V4 neurons to target stimuli in their corresponding receptive fields was enhanced after subthreshold microstimulation of FEF neurons that have corresponding retinotopy. Additionally, the activity of V4 neurons that had a receptive field that did not overlap with the movement field of the stimulated FEF neurons was suppressed. These results provide strong evidence that the FEF causally controls the gain of V4 responses.

Another approach that provides some causal evidence of prefrontal involvement in controlling the magnitude of responses to visual stimulation comes from event-related potential (ERP) studies in patients with lateral prefrontal damage. In a seminal study, Barcelo, Suwazono, and Knight (2000) presented patients with unilateral prefrontal cortex lesions with a visual attention task that asked patients to respond via button press whenever a target appeared in either visual field. It was observed that the amplitude of the P1 component of the ERP recorded from electrodes corresponding to visual areas was reduced for all stimuli presented to the contralesional visual field. Because this effect was independent of the relevance of the stimuli to the task, these results suggest that the PFC sends a tonic excitatory signal to ipsilateral extrastriate regions. Patients also exhibited attenuated target-related N2 and P3b responses suggesting both early and late aspects to attentional modulation.

More recently, attempts have been made to better characterize the relationship between the extrastriate cortex and a sources of top-down modulation by using transcranial magnetic stimulation (TMS) to perturb the function of frontal cortex while simultaneously recording activity in extrastriate cortex using both EEG and fMRI. In one of the first of such studies, Ruff and colleagues applied short trains of TMS to FEF while simultaneously recording fMRI activity from visual cortex (2006). The results of this stimulation had markedly different effects on visual cortex dependent on retinotopy. Increasing the intensity of the TMS pulses over FEF led to increased activity in the peripheral visual field, but suppressed activity in the central visual field in all retinotopic areas including V1-V4. An examination the behavioral effects of FEF stimulation showed that participants showed enhanced perceived contrast for peripheral relative to central visual stimuli, which is in accordance with what would be expected based upon the changes seen in visual cortex. Because these effects were only present when TMS was applied to FEF and not when applied to a control site, it can be concluded that FEF activity
causally modulate the gain in behaviorally relevant visual cortex activity.

Zanto and colleagues (2011) combined TMS with EEG in humans to examine the contribution of another frontal region in the gain modulation in early visual regions. Participants were instructed to attend to either the motion direction or the color of visual stimuli while ignoring the irrelevant feature. This information had to be maintained over a short delay for a subsequent memory probe. The authors took advantage of the fact that there are separate cortical regions known to process color (area V4) and motion (area V5). Each participant performed the selective attention task during fMRI scanning to find a PFC region, the right inferior frontal junction (IFJ), in each participant that showed increased functional connectivity with both V4 and V5 when the relevant feature was attended. In the next stage of the experiment, rTMS was targeted at this site just prior to the completion of the same task while EEG data were acquired. The amplitude of the P1 response in response to stimulus presentation has been shown to be sensitive to attentional modulation (Zanto and Gazzaley, 2009). There was an rTMS-related reduction in attentional modulation of the P1 amplitude to color stimuli and this decrease corresponded with a decrement in working memory accuracy. To investigate the mechanism of frontal interaction with visual cortical areas, the phase coherence of EEG oscillations between these two regions were examined. The alpha-band phase coherence between the IFJ and posterior visual regions just prior to stimulus onset was modulated by attention and this attentional modulation is likewise disrupted by frontal stimulation. These results not only provide strong evidence for the causal role of the IFJ in modulating the gain in activity of visual regions, but also provide insights into the manner in which these top-down signals are transmitted between disparate brain regions.

In another investigation of top-down control using a combined fMRI-TMS approach, Higo and colleagues (2011) used offline repetitive TMS (rTMS) to perturb function in an inferior PFC region just prior to the performance of a cognitive control task in the fMRI scanner. In this task, participants had to retain a subset of previously encoded information for a subsequent memory test. While activity in stimulus-selective regions of extrastriate cortex is typically enhanced when stimuli of the relevant category are the focus of attention and suppressed when they are ignored, this attention-dependent modulation was abolished following prefrontal disruption. These studies go beyond correlational and suggestive evidence for prefrontal control of the modulation of sensory cortex in favor of direct examinations of the causal source of top-down signals.
2.3 Selectivity of early visual regions

Different object categories are represented by spatially distributed yet overlapping regions in extrastriate visual cortex (Op de Beeck, et al. 2008; Haxby 2001). As previously mentioned, numerous imaging studies have suggested that the PFC can bias activity in these regions to enhance the processing of goal-relevant stimuli. To what extent does the PFC help maintain the selectivity of responses to different categories of stimuli in these visual areas? In one of the first “perturb and measure” approaches to studying top-down prefrontal modulation of sensory cortex, Fuster et al. (1985) cooled neurons in macaque PFC to look at its contribution to firing in inferior temporal cortex (ITC) and the successful performance of a delayed-match-to-sample task. During the delay period of the task, there was persistent stimulus-specific activity in ITC neurons. When PFC activity was attenuated via cortical cooling, the firing rate of ITC neurons during this delay period was reduced and the color-dependent differences in firing rate were diminished or abolished. These neuronal changes in spiking and stimulus-specificity coincided with reduced performance on the task. These results not only provide strong evidence for prefrontal involvement in top-down control signals, but they also provide clues as to the mechanism of this control. We can infer that the PFC both sends a gain signal back to ITC and that it has a hand in organizing assemblies of ITC neurons to react selectively to relevant stimuli.

As a follow-up to this study, Miller et al. (2011) sought to determine if a similar mechanism is present in humans. Using both patients with focal lesions to the PFC and rTMS in healthy normal participants, they were able to show that intact PFC function was essential for maintaining category selectivity in posterior visual regions. All participants completed a 1-back task in the scanner using faces and houses as visual stimuli. Both patients with frontal damage and participants receiving TMS to PFC prior to scanning showed less distinctiveness and greater similarity across extrastriate cortex between stimulus-evoked activity to face and scene stimuli in extrastriate cortex. By revealing that the PFC is essential for the tuning of category representations in extrastriate cortex, this work suggests that the ventral visual stream has a very dynamic structure and does not have a rigid representational structure. Both of these studies highlight a mechanism that would be virtually impossible to answer using recording techniques alone. Only by disrupting the prefrontal influence on ITC could it be determined if the category preferences in early visual areas are inherent to the tissue or dynamically tuned to the task at hand.
2.4 Timing of top-down influence

Another facet of top-down control that can perhaps best be ascertained by combining stimulation techniques with recording is the time course of signals emanating from PFC to posterior cortices. Resolving latencies in fMRI remains difficult due to the sluggishness of the hemodynamic response. Simultaneous multisite invasive recording is likewise technologically challenging. However, by comparing intact to disrupted PFC, the timing of responses can be contrasted to determine the time course of top-down signals. In a clever design, Tomita et al. (1999) severed posterior sections of the macaque corpus callosum in order to ensure that the ITC in each hemisphere could only receive visual information from the contralateral visual field and not through cross talk from the visual areas of the other hemisphere. During a memory task, the activity of neurons in ITC reflected remembered stimuli presented in the ipsilateral visual field despite the fact that they did not receive any bottom-up visual information. It is important to note that this “top-down” response had a latency ~200 ms longer than bottom-up responses suggesting that the visual information took a longer, more circuitous route. The authors hypothesized that the signal must have been sent via the PFC since the only intact corpus callosum fibers between the hemispheres were in anterior regions of the brain. This was confirmed by severing the remaining connections between the hemispheres, which abolished ITC activity in the “blind” hemisphere and disrupted task performance. These findings strongly support the idea that prefrontal signals bias processing in ITC and also suggests that this prefrontal involvement acts on the order of hundreds of milliseconds.

In a follow-up study to Barcelo et al. (2000), Yago and colleagues (2004) employed similar target-detection paradigm with patients with lateral PFC damage. This study improved upon the previous study by including a spatial attention component to the task that allowed cued participants to attend to only one visual field at a time. Using this new wrinkle in experimental design, the authors sought to disentangle the precise timing and nature of top-down signals from prefrontal regions to the relevant visual areas of the brain. The results replicated the previous finding of a tonic PFC signal to extrastriate cortex by showing that P1 ERP responses were reduced in both attended and unattended locations. Additionally, there were further deficits in target-related activity both early and late in the trial, but only at attended locations. These results provide strong evidence that the PFC not only exerts a tonic influence, but also has a phasic excitatory influence on extrastriate cortex to aid in the selection of relevant visual features in the environment.

In another study of the causal role of the FEF in visual attention, Taylor et al. (2007) applied brief, high-frequency trains of TMS to the right FEF during a spatial attention task while simultaneously recording ERPs from posterior electrode sites. Participants were asked to covertly direct their attention to either the left or right visual field based on cues given 500 ms prior to probe stimuli
requiring a button press if they appear in the attended hemifield. FEF stimulation (and not stimulation to a control site) during the cueing period of the task influenced visual ERPs measured from occipital cortex electrodes ipsilateral to the TMS site. This modulation of ERPs began immediately and was also observed following the presentation of the target stimuli that immediately followed. These changes in evoked activity are paired with detriments in behavioral performance, which illustrates the causal role of the FEF in modulating attention and elucidates the approximate timescale of its action on early visual regions.
How do the top-down signals from the frontal cortex differ with regard to the specific behavioral context in which a task is performed? Is the nature of the signal flexible, or is there a more tonic signal that is static across a variety of conditions? To answer these questions, Morishima and colleagues (2009) applied single TMS pulses over human PFC (electrode position FC2, which roughly corresponds to the FEF) while recording EEG activity in a selective attention task. Participants were asked to discriminate either the direction of the motion of a vertical grating or the gender of a face on the basis of a cue presented beforehand. By choosing to stimulate at varying time points during preparation, Morishima and colleagues were able to probe the system at different stages of cognitive processing to assay the functional relevance and time course of signals emanating from the PFC. Activity recorded from posterior electrode sites over the occipital and temporal lobes due to PFC TMS varied both as a function of stimulus type and depending on the amount of attentional preparation following the pre-stimulus cue. Additionally, these task-specific modulations in EEG activity were associated with differences in behavioral performance.

In another invasive study in monkeys, Ekstrom and colleagues (2008) paired the electrical microstimulation of FEF with fMRI recordings. FEF stimulation increased fMRI activity in higher-order visual areas that are known to be directly connected to the FEF in the absence of any visual stimulus. Interestingly, when stimulation was paired with a visual stimulus, enhanced activity was observed in the retinotopically corresponding regions of V1 and V2, which receive virtually no direct connections with FEF. This suggests that frontal control mechanisms can function in a state-dependent manner and that these top-down effects can be seen in brain regions that are multiple synapses away from a source of modulation.

A similar study combined TMS with fMRI to determine how prefrontal signals vary as a function of behavioral context (Feredoes et al. 2011). In this experiment, participants performed a delayed recognition task that required the retention of visual stimuli (either faces or houses) over a brief delay for a subsequent recognition probe. During the delay period, distracting visual stimuli from the opposite category to the memory targets were sometimes present. Previous work has shown that a region of the dorsolateral PFC (DLPFC) shows increased activity during delay periods containing distracting stimuli. While simultaneously recording whole brain BOLD responses with fMRI, short trains of TMS were applied to the right DLPFC during the delay period. DLPFC-TMS time locked to the presentation of distracting images led to activity increases in category-sensitive visual regions (FFA or PPA) representing the current memory targets. This effect was not observed when TMS was applied during delay periods without distracting stimuli. These results reveal that the mechanistic role of DLPFC is in protecting items in working memory from distraction and not simply...
in maintaining items over a delay.

2.6 Conclusion

Although there has been a large body of work providing suggestive evidence for the role of the PFC in top-down control operations, many of the most popular methods in neuroimaging and neurophysiology cannot provide causal evidence of this claim. In this review, we have highlighted several different approaches that can improve our understanding of the top-down modulatory influence of PFC by detecting the causal relationships between disparate brain regions. Only by combining stimulation, lesion, and recording methodologies can we overcome the limitations inherent to each technique and enhance our understanding of causal interactions within distributed cortical networks.
Chapter 3 The dynamic nature of top-down signals originating from prefrontal cortex: A combined fMRI-TMS study

3.1 Overview

The prefrontal cortex (PFC) has been proposed to be a source of top-down signals that can modulate extrastriate visual processing in accordance with behavioral goals. However, little direct causal evidence for this hypothesis exists. Using theta-burst transcranial magnetic stimulation, we disrupted PFC function in healthy participants before performance of a working memory task during fMRI scanning. PFC disruption decreased the category-specific tuning of extrastriate cortex responses, coinciding with decrements in working memory performance. We also found that activity in the homologous PFC region in the non-stimulated hemisphere predicted performance following disruption. Specifically, those participants with greater homologous PFC activity and greater connectivity between this region and extrastriate cortex were the most resistant to PFC disruption. These findings provide evidence for a compensatory mechanism following insults to the brain and insight into the dynamic nature of top-down signals originating from the PFC.

3.2 Background

Working memory refers to the temporary retention of information that was just experienced but no longer exists in the external environment, or was just retrieved from long-term memory (Baddeley, 1986; D'Esposito, 2007). Because of the capacity constraints of working memory (Cowan, 2005), task-relevant information must be effectively encoded for further maintenance by filtering out competing task-irrelevant information (Vogel et al. 2005). Numerous single-unit recording studies in nonhuman primates (Fuster and Alexander, 1971; Miller et al., 1996) and fMRI studies in humans (Curtis and D'Esposito, 2003) have demonstrated that the prefrontal cortex (PFC) and extrastriate cortex (EC) exhibit persistent activity during visual working memory delay tasks (Fuster and Jervey, 1982; Miyashita and Chang, 1988; Miller et al., 1993; Postle et al., 2003; Ranganath et al., 2004). Similarly, there is abundant evidence that neural activity is enhanced in those regions of EC that encode relevant stimuli (Fuster, 1990; Duncan et al., 1997; Kanwisher et al., 1998) and suppressed in regions that represent irrelevant stimuli (Kastner et al, 1998; Gazzaley et al., 2005). Based on this work, it is postulated that the PFC provides top-down signals to EC that selectively enhance the processing and active maintenance of goal-relevant sensory information (Knight et al., 1999; Miller and D'Esposito, 2005).

Although there are numerous studies that suggest the PFC interacts with EC to
subserve working memory, the vast majority offer only indirect evidence in support of this claim. Studies examining correlated neural activity between disparate brain regions cannot provide causal evidence that the PFC influences posterior brain regions. However, several lesion studies have provided direct causal evidence for functional interactions between the PFC and EC (Fuster et al., 1985; Tomita et al., 1999). In humans, event-related potential (ERP) recordings in patients with PFC lesions have shown that an intact PFC is needed for effective sensory processing in posterior visual areas during working memory tasks (Chao and Knight, 1998; Barceló et al., 2000).

Recently, attempts have been made to better characterize the relationship between the EC and the sources of top-down modulation by using transcranial magnetic stimulation (TMS) to perturb frontal cortex function while simultaneously or subsequently recording activity in EC using either EEG (Taylor et al., 2007; Morishima et al., 2008; Zanto et al., 2011) or fMRI (Ruff et al., 2006; Miller et al., 2011; Higo et al., 2011). While these studies have demonstrated that disruption of PFC activity can modulate the magnitude of activity in posterior visual areas, only Miller and colleagues (2011) have demonstrated that PFC disruption also modulates neural tuning. However, in this study, the relationship between behavioral performance and modulation of tuning could not be demonstrated.

The current study seeks to address the causal role of PFC-mediated top-down control of EC by administering TMS to disrupt PFC functioning prior to a working memory task performed during fMRI scanning. If the PFC is a source of top-down modulation, disruption of this region should alter the gain and the tuning of task-relevant representations within EC and lead to impaired performance on the task.
3.3 Materials and Methods

Participants
Twelve right-handed subjects (5 males, 7 females; age range, 18-29) participated in the experiment. All participants gave their written informed consent before participating in the study and received monetary compensation for their participation.

Experimental Timeline
Each participant underwent three separate fMRI acquisition sessions on separate days while performing the task. The first session was used to obtain a T1-weighted anatomical MRI scan and to acquire data to define regions of interest (ROIs) for use in subsequent analyses. Prior to the second and third sessions (counterbalanced across participants), TMS was applied either over a left inferior frontal gyrus (LIFG) region that was functionally defined from each participant’s initial scan, or a control brain region (somatosensory cortex). After each TMS session, participants were taken directly to the MR scanner. On average there was a nine-minute delay from the end of TMS and the acquisition of the first functional image.

Task Design
While being scanned during each of the three sessions, participants performed 5 runs of a modified N-back task each consisting of 4 blocks of 20 trials (Fig. 3.1). In this task, images of faces and scenes were sequentially presented and participants were instructed to selectively attend to and remember images from the relevant category while ignoring those from the irrelevant category, or to attend and remember both categories. At the beginning of each block, participants were given an instructional prompt to inform them which category of stimuli was to be remembered and which were to be ignored: 1) FACES - attend to faces and ignore scenes, 2) SCENES - attend to scenes and ignore faces, 3) BOTH - respond to both faces and scenes, 4) CATEGORIZE - indicate whether the current image is a face or a scene—with no attempt to remember them. On each trial, participants were presented with an image of either a face or a natural scene and had to indicate by button press whether or not the current image was a repeat of the image last seen that was of the same category. If an image was of the unattended category, participants had to simply press the “non-match” button. An identical number of faces and scenes were presented each in block and the number of “distracter” images from the unattended category that were presented between successive images of the attended category ranged between 0-3. Images were presented on the screen at fixation for 500 ms with a variable intertrial interval of 1.5s-4.5s. Because participants were instructed to maintain fixation throughout the experiment and previous work using inhibitory TMS to disrupt lateral PFC has failed to produce impairments in simple perceptual judgments (Rounis et al. 2010), impaired task performance should reflect
impaired top-down processing and not low-level visual interference.

Figure 3.1 Structure of the behavioral task. Participants were prompted at the beginning of a block which stimulus categories to remember (FACES, SCENES, BOTH) or identify (CATEGORIZE). In the ATTEND FACES, ATTEND SCENES, and ATTEND BOTH conditions, on each trial, participants manually responded whether or not the image currently on the screen was a repeat of the last image of the relevant stimulus category. Each image needed to be maintained across all intertrial intervals (1.5-4.5 s) and intervening distracter images (up to three). In this schematic illustration of the task, a box surrounding a stimulus indicates that it is from the category that needed to be remembered; these boxes were not presented to the participant. In the CATEGORIZE condition, participants simply responded whether the current image was a face or a scene.

Transcranial Magnetic Stimulation

For all experiments, participants were seated in a comfortable chair. Electromyography was recorded using electrodes from the right first dorsal interosseous (FDI) muscle of the dominant hand in all participants. TMS was applied using a hand-held figure-eight coil with an outer winding diameter of 70
mm (MagstimCo., Whitland, Dyfed, UK). All pulses were delivered using a Magstim rapid stimulator connected to 4 booster modules that produce biphasic pulses. Stimulation was delivered over the motor cortical hand area with the coil placed tangentially to the scalp with the handle pointing posteriorly. The motor cortical hand area was defined as the location on the scalp where magnetic stimulation produced the largest motor-evoked potential (MEP) from the contralateral FDI when the participant was relaxed. The stimulation intensity was defined in relation to the active motor threshold (AMT) of the participant. The AMT was defined as the minimum intensity of single-pulse stimulation required to produce an MEP on more than 5 out of 10 trials from the contralateral FDI while the participant was maintaining a voluntary contraction of about 20% of maximum in the FDI. Raw EMG signal on the screen was provided as visual feedback to the participant to help maintain a constant muscle contraction of the correct force.

The stimulation targets were localized using Brainsight, a computerized frameless stereotaxy system (Rogue Research, Montreal, Canada). This system uses an infrared camera to monitor the positions of reflective markers attached to the participant’s head. Head locations are related to the participant’s previously acquired structural MRI data in real time after the data have been co-registered to a set of anatomical locations. Reflective markers are attached to the coil and the participant so that relative positions of the coil to the head (and the MRI) can be tracked, allowing precise positioning of the coil with respect to previously chosen MRI locations.

Many repetitive TMS (rTMS) studies have used 1 Hz stimulation to cause transient reductions in cortical excitability that outlast the period of stimulation. However, the reported effects of this TMS protocol are short lasting (usually 15 minutes or less) and often weak and highly variable (Gangitano et al., 2002). In contrast, continuous theta burst TMS is a more recent TMS protocol developed to address this shortcoming (Huang et al., 2005). Continuous theta burst TMS parameters used in the current study were identical to those described by Huang and colleagues (2005). Theta burst TMS consists of 50 Hz trains of three TMS pulses continuously repeated every 200 ms over a period of 40 seconds (600 pulses total) and has been shown to depress activity in the stimulated region for up to 60 minutes following stimulation. This longer duration of altered excitability is ideally suited for studying the effects of TMS with fMRI given that many imaging studies require repeated acquisition runs lasting over 30 minutes.

**Stimulation Sites**

The left inferior frontal gyrus (LIFG) was chosen for TMS because it is a likely source of top-down modulation of early visual areas, based on functional connectivity analyses we have performed on data collected during performance of the cognitive task used in the current study (unpublished data). While many studies have opted to use the vertex of the skull as a control region for TMS, it
often falls between cerebral hemispheres and is only likely to mimic the sensation of stimulation and not actually stimulate brain tissue. Thus, to control for changes due to non-specific stimulation of brain tissue as well as the scalp sensation of stimulation, the control region used in this study was the left postcentral gyrus. This site was not activated by our task in any participant and it is not directly anatomically connected to posterior visual areas.

**Functional MRI Acquisition and Pre-processing**

MR data were acquired with a Siemens 3 Tesla scanner (Berlin/Munich, Germany) with a 12-channel head coil. Functional data were obtained using a 2-shot $T_2$-weighted echo-planar imaging (EPI) sequence sensitive to blood oxygenation level-dependent (BOLD) contrast ($TR = 1000$ msec, $TE = 32$ msec, 230 mm field of view with a 64 x 64 matrix size, in-plane resolution 3.5mm x 3.5mm). Each functional volume contained 18 contiguous 3 mm-thick axial slices separated by a 0.5 mm interslice gap. Whole-brain MP Flash $T_1$-weighted scans were acquired for anatomical localization. Functional data were realigned to the first volume acquired and spatially smoothed with a 4-mm full-width at half-maximum Gaussian kernel.

**Behavioral Analysis**

Mean response times (RTs) and accuracy rates were computed for each participant for all of the conditions (BOTH, FACES, SCENES, CATEGORIZE). The effect of TMS on the task was evaluated by a repeated measures analysis of variance (ANOVA) on the RTs and accuracy with the stimulation site (control and LIFG) and the condition as within-subject factors. Two-tailed t-tests were used to compare results between the two stimulation conditions where appropriate.

**Univariate fMRI Analysis**

Task-dependent changes in the BOLD signal were modeled with independent regressors for each attention condition (BOTH, FACES, SCENES, or CATEGORIZE) and each stimulus type (face or scene). These regressors were generated by convolving a gamma function with a vector containing the onset times for each trial type. Trials with incorrect behavioral responses were modeled separately and excluded so that the resulting analyses only considered trials in which participants made a correct response. Statistical contrasts indexing attentional enhancement for each condition were computed as weighted sums of the estimated beta coefficients (ATTEND FACES = attended face stimuli vs. ignored face stimuli; ATTEND SCENES = attended scene stimuli vs. ignored scene stimuli; ATTEND BOTH = All BOTH stimuli vs. all CATEGORIZE stimuli). Each run was mean-centered and detrended for linear and polynomial trends using linear least squares. Maps of the parameter estimates (beta values) were computed from the GLM from each session and normalized on a subject-by-subject basis to each individual's anatomical scan. Mean parameter estimates
for each regressor were also calculated within each functionally defined ROI for each subject. Two-tailed, paired sample t-tests were used to test for the significance of differences across groups and across task contrasts for each ROI.

**Regions of Interest (ROIs)**

Left inferior frontal gyrus (LIFG), fusiform face area (FFA), and parahippocampal place area (PPA) ROIs were functionally defined using different contrasts from each individual’s first session. LIFG ROIs were defined as the most statistically significant cluster of activity in the LIFG based on a contrast of all attended images versus all ignored images (Fig. 3.2). Bilateral FFA ROIs were created by selecting the cluster of activity in the fusiform gyrus of each hemisphere that exhibited the strongest preference for faces versus scenes during the task, as assessed by a t-test. Similarly, bilateral PPA ROIs were created by selecting the cluster of voxels in the parahippocampal/lingual gyrus that exhibited the strongest preference for scenes vs. faces. Right Inferior Frontal Gyrus (RIFG) ROIs were created by taking the mirror-flip of each individual’s functionally defined LIFG ROI. Left postcentral gyrus (PCG) ROIs were anatomically defined with reference to the Duvernoy brain atlas (1999). Each participant’s left PCG was identified on the anatomical scan and ROIs were drawn as spheres with a radius of 5mm centered 10mm away from the midline and 5mm from the top edge of the brain.

![Figure 3.2 LIFG TMS site across participants. ROIs are](image)

**Figure 3.2** LIFG TMS site across participants. ROIs are...
superimposed on axial slices of an average of all participants’ structural MRI scans normalized to MNI space (z coordinate is shown for reference). Each diamond represents the site for one participant.

Functional Connectivity Analysis

Functional connectivity maps were created on a subject-by-subject basis using a beta-series connectivity analysis previously described by Rissman and colleagues (2004). This approach allows us to determine correlations on a trial-by-trial basis between brain regions as opposed to blocks of trials. This analysis produces a whole brain map of Pearson’s r values for each subject that were subsequently transformed using a Fisher’s r-to-z transformation.

Category Selectivity Analysis (Spatial Correlations)

Previous work has shown that while EC seems to have distinct modules that appear to process certain categories of visual stimuli preferentially (e.g. faces and scenes), there is evidence that the representation of this information is actually quite widely distributed and overlapping (Haxby, et al. 2001; Op de Beeck, et al. 2008). Can the PFC influence which regions of cortex show selectivity and can these top-down signals enhance the clarity of processing across EC by biasing activity toward a preferred stimulus category? To determine the distinctiveness of scene and face representations in EC we employed a spatial correlation analysis. In this analysis, the similarity of patterns of neural activity evoked by the task after TMS to the two different stimulation sites was assessed. Calculating the correlation between spatial patterns of activity evoked by the task under the different stimulation conditions offers another method for evaluating the change in neural activity as a result of frontal stimulation since spatial correlations are relatively insensitive to differences in the magnitude of evoked activity that might occur between the two stimulation conditions.

In this analysis, large functional ROIs comprised of the union of all scene-selective and face-selective voxels across all of EC (including, but not limited to, FFA and PPA) during each participant’s initial scan were defined as each participant’s left and right EC. These ROIs were subdivided into a set of new ROIs comprising the most task-active voxels (top 30, 40, 50, 60, 70, 80, 90, 100, 110, and 120 voxels) as defined by a contrast of all attended images versus all ignored images. For each stimulation condition (IFG or control), the beta value for the corresponding covariate was obtained for each voxel within each ROI. These values were translated into a linear vector for each ROI for each condition (Aguirre, 2007). Correlations (Kendall’s tau) of the evoked activity between the two stimulus categories (face and scene) within each task were calculated. These correlations are the spatial correlations of the pattern of activity evoked by faces versus scenes. For each participant, the difference in the correlation value
between IFG stimulation and control stimulation was calculated and used as an index of the change in category selectivity due to frontal stimulation.

The significance of the observed difference in correlations was assessed by creating a simulated distribution of correlation values via bootstrapping by randomizing the values within each subset of voxels before calculating each correlation. Shuffling was done 10,000 times per subject for each stimulation condition. A difference score was calculated for each bootstrapping iteration and this resulted in a Gaussian distribution of difference values. Difference values more extreme than 95% of the distribution were taken to be significant.

**Brain-Behavior Correlation Analysis**

Brain-behavior correlation plots were created by plotting the mean t-value or z value (from the connectivity analysis) from a given ROI for each participant as a function of that participant’s mean RT or accuracy for the given contrast of interest. Correlations between brain activity and behavior were quantified with Pearson’s coefficient $r$ and Spearman’s rank-order coefficient $r_s$. 
3.4 Results

Behavioral analyses

A repeated measures ANOVA of the accuracy and response times (RT) of responses during the task revealed a main effect of condition (accuracy - \( F_{1,11} = 11.74, P < 0.01 \); RT - \( F_{1,11} = 34.17, P < 0.001 \)) and distracter number (accuracy - \( F_{1,11} = 8.71, P = 0.01 \); RT - \( F_{1,11} = 16.10, P < 0.005 \)). Participants had slowest RT during the ATTEND BOTH condition, followed by the ATTEND FACES and ATTEND SCENES condition, then the CATEGORIZE condition. When averaged across stimulation conditions, participants also had the worst accuracy in the ATTEND BOTH condition, but this seems to be driven by especially poor accuracy following LIFG TMS. The linear contrast for the TMS site x condition interaction was also significant (accuracy - \( F_{1,11} = 5.87, P < 0.04 \)). As predicted, performance suffered after LIFG stimulation when compared to control stimulation (Fig. 3.3), but only during the ATTEND BOTH (\( t_{11} = 2.41, P < 0.05 \)) and ATTEND FACES condition (\( t_{11} = 2.35, P < 0.05 \)). There was also a three-way TMS site x condition x distracter interaction in the response times (RTs) (\( F_{1,11} = 10.64, P < 0.01 \)). This interaction was driven by slower RTs in the ATTEND SCENES condition with three distracters after LIFG stimulation when compared to control stimulation (\( t_{11} = 1.70, P < 0.06 \)). Accuracy in responding “non-match” to irrelevant distracter stimuli (faces in the ATTEND SCENES condition and scenes in the ATTEND FACES condition) was virtually identical and near ceiling across the TMS conditions (LIFG TMS – 98.3%, Control TMS – 98.7%; \( t_{11} = 0.68, P = 0.52 \)) suggesting that participants were able to keep track of the task rule following PFC stimulation.

Univariate fMRI analyses

Consistent with our previous work (Gazzaley et al., 2005), all participants showed enhanced FFA activity when face images were attended (i.e. relevant) versus when they were ignored (i.e. irrelevant) and enhanced PPA activity when scene images were attended versus ignored. The magnitude of this gain effect during trials in which stimuli were relevant vs. irrelevant can be considered an index of top-down modulation. Thus, we examined the effect of LIFG on this metric within the FFA and PPA. In the two behavioral conditions in which LIFG TMS, as compared to control TMS, caused worsened performance (ATTEND FACES condition and ATTEND BOTH condition), FFA and PPA activity, as indexed by contrasts reflecting attentional enhancement to relevant images, was reduced after LIFG TMS. PPA activity in the ATTEND SCENES condition was not reduced following LIFG TMS. None of these effects, however, were significant at the group level, likely due to significant variability across subjects. This individual variability in the BOLD data prompted us to determine whether it could be accounted for by individual differences in the effect of TMS on behavioral performance.
Behavioral results. Accuracy during the ATTEND FACES and ATTEND BOTH conditions of the task was reduced after TMS to LIFG as compared to TMS of a control region. Data are represented as mean +/- SEM. * denotes P < 0.05.

Brain-Behavior Correlation Results

To further explore the relationship between TMS effects on brain activity and behavioral performance, BOLD responses in each frontal and extrastriate ROI (i.e. left and right IFG, left and right FFA) for each contrast were correlated with participants’ behavior during the conditions in which behavioral performance was affected by TMS (i.e. ATTEND FACES and ATTEND BOTH conditions).

Following TMS to the control region, LIFG was the only ROI that exhibited a significant relationship between brain activity during the ATTEND FACES condition and behavior (RT and accuracy). Specifically, there was a significant positive correlation between performance and the magnitude of activity in LIFG in response to attended face stimuli. That is, the most accurate participants exhibited the greatest increase in BOLD signal for this contrast ($r_s = 0.79$, $P < 0.01$; Fig. 3.4). Consistent with our prediction that the LIFG is a source of top-down signals modulating the gain of EC, this brain-behavior relationship during the ATTEND FACES condition was not significant after LIFG TMS ($r_s = 0.27$, $P = 0.40$). Moreover, these relationships were significantly different between control
and LIFG TMS ($z = 1.69$, $P < 0.05$ one-tailed). There were no significant correlations between behavior and BOLD signal for any region during the ATTEND BOTH or ATTEND SCENES condition.

**Figure 3.4** Correlation plots of IFG activity vs. Task Accuracy during the ATTEND FACES condition. Following TMS to the control region, LIFG activity, but not RIFG activity correlated with accuracy. Following TMS to LIFG, the opposite pattern was observed — LIFG activity no longer correlated with accuracy whereas RIFG activity and accuracy became correlated.

To explore possible compensatory mechanisms supported by the non-stimulated hemisphere, we also examined the relationship between behavior and BOLD signal in the right IFG (RIFG). While there was no relationship between RIFG activity and performance in the ATTEND FACES condition after control stimulation ($r_s = 0.25$, $p = 0.43$), activity in this region was positively correlated with accuracy following LIFG stimulation ($r_s = 0.73$, $P < 0.01$; see **Fig. 3.4**). To further investigate the contribution of RIFG function to successful performance following LIFG TMS, we assessed the relationship between functional connectivity between RIFG and right FFA and the TMS-induced change in accuracy during the ATTEND FACES condition. A change in accuracy was calculated as the difference between performance after control TMS versus after PFC TMS (control – LIFG TMS). Functional connectivity between the RIFG and the right FFA after control stimulation successfully predicted this change in accuracy following LIFG TMS ($r_s = 0.545$, $P < 0.07$). Furthermore, a median split of the participants based on the change in performance between LIFG TMS and control TMS showed that poor performers had significantly less connectivity.
between the RIFG and the right FFA after Control TMS (t_6 = 2.594, P < 0.05; Fig. 3.5) These findings suggest that preserved behavioral performance following LIFG disruption depends on the ability of the non-stimulated right hemisphere to compensate for altered LIFG function.

These findings suggest that preserved behavioral performance following LIFG disruption depends on the ability of the non-stimulated right hemisphere to compensate for altered LIFG function.

Figure 3.5  RIFG-RFFA connectivity and TMS-related change in accuracy during the ATTEND FACES condition. Participants who maintain high levels of task performance after LIFG TMS were those who had high levels of RIFG-RFFA connectivity after control TMS.

Spatial Correlation Results

The above results suggest that PFC can modulate the gain of activity within EC. However we also sought evidence that PFC can modulate the tuning of EC representations. If true, LIFG TMS should lead to decreased category selectivity which would manifest as higher spatial correlations between stimulus-evoked activity within EC in response to attended images of one category (face or scene) and ignored images of the other category.

As predicted, LIFG TMS, as compared to control TMS, significantly increased the spatial correlation between the patterns of activity in EC evoked by attended faces and ignored scenes in the ATTEND FACES (P < 0.05 via permutation testing). These increases were observed in both left and right EC regardless of how many voxels were included in the analysis (within a range of 30-120 voxels, Fig. 3.6).
Figure 3.6 Category selectivity (as measured by Kendall’s Tau) in bilateral extrastriate cortex between attended faces and ignored scenes during the ATTEND FACES condition of the task. TMS to the LIFG led to increases in spatial correlation between evoked activity to the face and scene stimuli, which indicates a decrease in category selectivity in the underlying cortex. Error bars represent s.e.m.

To assess how this change in selectivity to face and scene stimuli following LIFG TMS affected working memory performance, we plotted the change in accuracy in the ATTEND FACES condition for each participant who exhibited a decrement in performance following LIFG TMS against their change in spatial correlation. Indices of the TMS-induced change in accuracy and spatial correlation were calculated as the difference between LIFG and control TMS. As predicted, the greater the increase in spatial correlation between extrastriate activity evoked by faces and scenes as a result of LIFG TMS, the greater the decrement in accuracy exhibited by participants. Again, when averaged across hemispheres this relationship was significant across a range of ROI sizes (80-120 voxels, $r = 0.76-0.93$, $P < 0.05$; Fig. 3.7). Moreover, it was also significant when the stimulated hemisphere was examined alone (within a range of 30-110 voxels, $r = -0.72-0.84$, $P < 0.05$) and approached significance in the non-stimulated hemisphere.
Figure 3.7 Correlation plot between TMS-induced change in category selectivity (as measured by Kendall's Tau) in bilateral extrastriate cortex (120 voxel ROI size) and change in accuracy during in the ATTEND FACES condition. Following TMS to LIFG, an increase in spatial correlations (indicative of a decrement in category selectivity) is predictive of a greater decrement in accuracy on the task. $\Delta = \text{IFG TMS} - \text{Control TMS}$. 

$r = -.93, p < .001$
3.5 Discussion

To assess the causal influence of the PFC on top-down modulation of posterior visual processing regions, we used continuous theta burst TMS to disrupt function in the LIFG prior to the performance of a working memory task. Our results show that after stimulation, LIFG function was altered such that activity in this region was no longer predictive of performance on a working memory task. We also show that this frontal disruption leads to a change in the selectivity of EC and that this change corresponds with a decrement in behavioral performance. Our data also suggest that those participants who are best able to maintain a high level of performance after LIFG stimulation recruit the homologous RIFG as a possible compensation for disrupted function in the stimulated hemisphere. Additionally, the participants who had greater connectivity between the RIFG and the right FFA prior to PFC TMS were the most resistant to disruption, that is, they exhibited the least decrement in behavioral performance.

It must be noted that while the relationship between BOLD activity and behavior changed following LIFG TMS, the overall magnitude of the response in this region did not differ significantly between sessions. While there have been many studies combining inhibitory TMS protocols with neuroimaging, the effect of stimulation on the BOLD signal under the coil remains unclear (Sack et al. 2007; Bestmann et al 2008; Hubl et al 2008). Similarly the mechanism by which theta-burst stimulation acts on neural activity remains elusive, although there has been some evidence that suggests its similarity to long-term depression (Huang et al. 2007). Nevertheless, there are several potential explanations for how neural activity could be disrupted without a concomitant change in BOLD activity. It is possible that population-level neuronal activity remains at a similar level but that the specific patterns of activity are inefficient or inadequate for performance on the task at hand. It is also unlikely that all neuronal firing would be suppressed in a stimulated region. There is evidence from studies examining patients with brain damage that there is often compensatory activity in perilesional tissue (Nudo 1999). After TMS, unaffected tissue adjacent to the disrupted tissue might exhibit increased activity to compensate for the loss of function. Unfortunately, with the relatively coarse spatial resolution of fMRI and the ROI approach that we have taken, it is not possible to ascertain which portions of left IFG are disrupted and which, if any, are unaffected and/or compensating.

While studies using functional connectivity analyses to examine network interactions during top-down modulation are informative (Miller and D'Esposito, 2005), these methods are correlational and do not provide strong evidence for causal interactions. Recently, two separate studies have shown that the frontal eye fields (FEF) play a causal role in attentional processing by combining TMS with EEG (Taylor et al., 2007; Morishima et al., 2008). Both studies found that perturbing function in the FEF leads to decreased task-evoked activity at posterior electrode sites. A combined TMS-EEG study extended these findings by stimulating another frontal control region, the inferior frontal junction (IFJ), and
observing similar gain changes in activity in electrodes recording activity in early visual areas (Zanto et al., 2011). By combining TMS with fMRI, we replicated and extend these findings by demonstrating that top-down signals emanating from PFC act via both gain and tuning mechanisms. Moreover, the better spatial resolution of fMRI allowed us to determine that the PFC in a homologous region in the non-stimulated hemisphere can provide compensatory top-down signals following PFC disruption in the opposite hemisphere. Recently, another study used TMS to disrupt PFC function prior to the performance of a selective attention task performed during fMRI (Higo et al., 2011). In this study, PFC TMS diminished top-down modulation of EC but did not impact the bottom-up activation of these same regions. However, these TMS-induced activity changes were not associated with changes in behavioral performance. In contrast, our results demonstrate that the magnitude of gain and tuning changes in EC following PFC disruption is predictive of subsequent behavioral performance on a working memory task.

In addition to providing evidence that top-down signals from PFC can modulate the gain of activity in early visual areas, we also sought to determine if top-down signals could modulate neural tuning. To investigate tuning, we measured the selectivity of BOLD responses in early visual areas to different categories of visual stimuli (i.e. faces vs. scenes). Previous work in nonhuman primates has shown that neurons in inferotemporal cortex (ITC) respond preferentially to specific colors during a delayed match-to-sample task and that disrupting PFC function can lead to decrements in behavioral performance (Fuster et al., 1985). This decrement in performance coincided with a disruption of the color-dependent differences in firing rate among ITC neurons such that these neurons abandoned their previous color preference and treated different colored samples similarly. In another fMRI study (Miller et al., 2011), we demonstrated that disrupting PFC function can reduce the selectivity of responses to face and scene stimuli in EC, but this change in tuning could not be linked to behavioral performance. The current study replicates this finding but also demonstrates for the first time in humans that altered tuning coincides with disrupted behavioral performance on a working memory task.

Studies of patients with brain lesions have shown that working memory function is left relatively unimpaired following unilateral PFC damage (D’Esposito and Postle, 1999). Recent work suggests that increased activity in the intact hemisphere in these patients may serve as a compensatory response for the loss of function in damaged tissue (Corbetta et al., 2005; Nudo, 2007; Voytek et al., 2010). Consistent with these findings, we found that only after TMS to LIFG, activity within the homologous RIFG predicted successful performance on the working memory task. A recent study of patients with diffuse axonal injury due to brain trauma reported a similar finding (Turner et al., 2011). Control participants activated a left-lateralized PFC network that correlated with working memory function, whereas homologous right PFC activity was not. However, in patients right PFC activity did correlate with working memory function. The authors
attributed this pattern of findings to “unmasking of extant, but behaviorally latent” brain networks supporting working memory rather than functional reorganization per se (i.e. recruitment of brain networks not engaged by controls on the task). We believe that a similar mechanism also explains our findings.

Other patient studies have suggested that there are transcallosal interactions between homologues in the two hemispheres, which are likely inhibitory, and that unilateral damage (or TMS disruption) can cause increased excitation of the undamaged hemisphere (Sprague, 1966; Kapur, 1996; Shimizu et al., 2002; Kobayashi, et al., 2004). Thus, studies of patients with focal lesions that have demonstrated that TMS disruption of the contralesional hemisphere ameliorate behavioral deficits may do so by releasing the perilesional tissue from transcallosal inhibition (Oliveri et al., 1999; Mansur et al., 2005; Naeser et al., 2011). In our study, we propose that TMS disruption of LIFG function may have led to increased activity in the homologous RIFG leading to a release of a dormant (latent) network that could support working memory function. However, the time scale of the effects we observed is significantly different from those observed in patients with chronic lesions from either stroke or traumatic brain injury. Prior studies using electrophysiological recordings in non-human primates have demonstrated that PFC neurons are highly adaptable and can code for newly learned stimulus categories (Freedman et al., 2001), learn new stimulus-response associations over a matter of trials (Asaad et al., 1998), and even represent two different types of information over different delay periods of a single trial of a working memory task (Rao et al., 1997). The compensatory effects we observed immediately after LIFG disruption are consistent with these prior findings and provide further support that PFC function can adapt on a relatively short time scale in response to behavioral demands.
Chapter 4. Disruption of frontal cortex function improves unconscious perceptual memory

4.1 Overview

Attentive encoding has frequently been shown to correspond with accurate responses in recognition tests. Recently, a pair of studies described conditions under which recognition memory for kaleidoscope images is improved under divided attention conditions designed specifically to distract participants at encoding (Voss et al., 2008; Voss and Paller, 2009). These studies suggest that explicit memory processes in the prefrontal cortex (PFC) at encoding might be harmful to subsequent recognition accuracy. However, the methods used in this work lack the ability to assess whether the PFC is necessary for this attentive encoding and could not provide evidence for which regions of the PFC are involved in this task.

The present study administered continuous theta burst (cTBS) repetitive transcranial magnetic stimulation (rTMS) over two sites in the PFC just prior to encoding in order to examine the effects of frontally-mediated attention on memory retrieval. If the PFC is a source of harmful attention effects on encoding, then disruption of function in these regions should lead to an increase in recognition accuracy. The results suggest that after stimulation of the PFC, recognition accuracy is improved relative to recognition accuracy after a control TMS condition. Crucially, this improved performance occurs only in responses in which the participant's awareness of retrieval was absent. This finding indicates that attentional processing in the PFC may negatively impact successful recognition memory by interfering with implicit memory processes at encoding and at test.
4.2 Background

A large body of work in cognitive psychology, neuropsychology, and cognitive neuroscience over the last 30 years has focused on the distinctions between explicit memory and implicit memory (Gabrieli, 1998; Squire, 2004; Squire, 2009). Explicit memory (sometimes referred to as declarative memory) is usually examined by using tests of free recall and recognition and typically refers to the conscious intentional retrieval of past information. Implicit memory (or non-declarative memory), in contrast, refers to memory that unconsciously drives behavior and is usually uncovered by using tests of priming whereby prior exposure facilitates further performance on tasks using the repeated material.

Several influential studies reporting that only explicit memory, and not implicit memory, is impaired in patients with amnesia has led to the dominant view in the literature that implicit and explicit memory rely on the activity of distinct brain networks (Cohen and Squire, 1980; Graf and Schacter, 1985; Gabrieli et al., 1995). These patient studies have been followed up by numerous behavioral, electrophysiological, and neuroimaging studies that have provided further evidence that explicit and implicit tasks dissociate behaviorally and in neural activation (Gabrieli, 1998; Squire, 2004). In studies of perceptual learning, explicit memory has been shown to depend primarily on the control processes of medial temporal lobe structures such as the hippocampus (Schacter and Wagner, 1999) and cortical regions such as the prefrontal cortex (Blumenfeld and Ranganath, 2007). In contrast, implicit perceptual memory is thought to be supported by the fluency of neural processing in the cortical networks involved in the initial perception (Wiggs and Martin, 1998).

However, while these implicit and explicit forms of memory seem to be separable in terms of their neuroanatomical underpinnings, the interaction between these two systems remains unclear (for a review, see Dew and Cabeza, 2011). Can implicit memory processes influence explicit memory processes and vice versa? Recently, a pair of studies described conditions under which taxing explicit memory resources by attentional distraction can improve perceptual recognition memory without awareness (Voss et al., 2008; Voss and Paller, 2009). During recognition tests in these two experiments, participants were highly accurate in identifying old items when they expressed no confidence in their responses and claimed that they were guessing randomly. Because dividing attention at encoding typically disrupts explicit memory processes (Mulligan, 1998), these results suggest not only that implicit memory can aid performance on recognition tests but also that explicit processing can interfere with the expression of this implicit memory. Furthermore, event-related potentials recorded during the task suggested that explicit recognition (denoted by “remember” and “know” responses by the participants) coincided with increased activity in the frontal cortex while implicit recognition (“guess” responses) did not (Voss and Paller, 2009).
While patients with damage to the medial temporal lobe often have profound anterograde amnesia (Squire et al., 2004), memory deficits in patients with frontal lobe damage are more subtle (Milner 1962; Shimamura 1995; Ranganath and Knight 2002). Additionally, neuroimaging studies of long-term memory have consistently shown that higher levels of activity in the prefrontal cortex (PFC) during encoding leads to greater accuracy during later tests of recall or recognition (Paller and Wagner, 2002). Because of this work, the frontal cortex’s contribution to long-term memory performance is thought to be the selection and organization of to-be-remembered information at encoding and retrieval (for a review, see Blumenfeld and Ranganath, 2007).

Can explicit memory processes in the frontal cortex interfere with the expression of accurate implicit memories? In the work of Voss and Paller (2008,2009), attentional distraction is used to deplete the explicit resources of the PFC during encoding. A more direct test of whether PFC processes are responsible for interfering with accurate implicit memory responses is to directly perturb function in the frontal cortex. Here we used continuous theta-burst transcranial magnetic stimulation (TMS) to disrupt the function of two distinct regions of the PFC in humans prior to performance of a perceptual recognition memory task adapted from Voss and Paller (2009). If activity in these regions reflects explicit memory processing, disruption of this activity should lead to superior recognition memory without awareness.
4.3 Materials and Methods

Participants
Twenty-six right-handed subjects (10 males, 16 females; age range, 18-29) participated in the main experiment. A mixed design was used with in which participants were randomly assigned to one of two groups. Each participant underwent two separate experimental sessions with TMS on separate days. Participants in the ventrolateral prefrontal cortex (VLPFC) group (13 participants) received stimulation to VLPFC in one session and stimulation to the vertex of the scalp in a separate session to control for any nonspecific secondary effects of TMS. Likewise, participants in the dorsolateral prefrontal cortex (DLPFC) group (13 participants) received stimulation to DLPFC in one session and to vertex in another. The order of stimulation (frontal vs. vertex) was counterbalanced across participants in each group. Additionally, age, years of education and gender were matched between groups. All participants gave written informed consent before participating in the study and received monetary compensation for their participation.

Stimuli
Visual stimuli consisted of 14 sets of 12 perceptually paired kaleidoscope images (336 images total). They were created by overlaying three opaque hexagons of different color and bisecting and randomly deflecting the sides of each three times. All of the images in each set used the same three hexagon colors. Perceptual matching for each pair was achieved by using the same colors, employing similar deflection angles, and visual inspection by a research assistant.

Task Design
Following the administration of TMS, kaleidoscope images were presented to individuals in 5 separate study-test blocks (Fig. 4.1). During the encoding period of each block, 12 target images were presented for 200 ms each with a 1.5 s interstimulus interval. Targets were chosen randomly from each perceptually matched pair of images for each participant. A forced-choice recognition test was administered after a 45 s delay upon the conclusion of the encoding period. Each test trial consisted of one of the 12 targets studied during encoding and a perceptually matched foil alternately presented 3 times for 500 ms each with a 1 s interstimulus interval. Targets were randomly assigned to be in the first or second interval in the test trial. Participants were instructed to visually compare the two stimuli and to indicate which item was previously studied using a button press.

Immediately following each recognition decision, participants were presented with a metamemory prompt that asked them to report their awareness of memory
retrieval using a remember/know procedure. Participants were instructed to give a remember response if they were able to retrieve a specific detail about the item or the context in which they saw the item during study. Know responses were given if they were confident about their response but could not recollect any specific details from encoding but they had a sense of familiarity for the selected image. Guess responses indicated a lack any feelings of familiarity and participants feeling that they were just “guessing randomly” because they needed to make a response. In order to ensure that guess responses were relatively frequent, participants were told that guess responses would likely be most frequent and that they should not hesitate if they felt that they were guessing (Jeneson et al., 2010; Voss and Paller, 2010). Before the experimental blocks, participants practiced the task with a mini-block that consisted of only 3 study items and a corresponding forced-choice recognition test.

![Experimental design](image)

**Figure 4.1** Experimental design. After viewing each set of images, participants were given a forced-choice recognition test for each image presented at study. Following each response, a metamemory judgment was made on the subjective “feeling of knowing”.

**Frontal Stimulation Targets**

Both frontal cortex TMS targets were chosen based on previously reported activations from functional magnetic resonance imaging (fMRI) studies using standard-space coordinates from the Montreal Neurological Institute (MNI) brain. The MNI-space coordinate for VLPFC stimulation was based primarily on a coordinate published in Wagner et al (1998). This study found that activity in a
VLPFC region, the inferior frontal gyrus (IFG; centered on -51, 25, 12), during encoding correlated with subsequent explicit item recognition. In the present study, we used a similar coordinate that was situated more clearly on the cortical surface of pars triangularis (-53, 28, 12). This location was determined to be more reliable landmark for frameless stereotaxy. The coordinate for the DLPFC stimulation site was based on a local maximum from a study by Blumenfeld et al (2011). This study found that a region of DLPFC, the middle frontal gyrus (MFG; centered on -44, 35, 26), correlated with subsequent explicit long-term memory for inter-item relational information. In the present study, we used a similar coordinate that was positioned slightly more dorsal (-43, 35, 30). It should be noted that although the VLPFC and DLPFC stimulation sites are on adjoining gyri, care was taken to ensure that the coordinates that were chosen were at least 20mm apart in every direction to ensure that stimulation did not bleed over from one site to the other. 10mm spherical masks were constructed at the VLPFC and DLPFC stimulation sites in MNI-space and reversed normalized into individual subject space using the high-resolution structural MR image acquired for each participant during prior studies (Fig. 4.2). These masks as well as the structural MR images were used for frameless stereotaxy (see below).

**Figure 4.2** Theta-burst TMS targets. MNI-space coordinates for both ventrolateral and dorsolateral PFC regions important for subsequent memory were taken from prior fMRI studies and reverse-normalized into each subject's native space structural MRI scan.
For all experiments, participants were seated in a comfortable chair. Electromyography was recorded using electrodes from the right first dorsal interosseous (FDI) muscle of the dominant hand in all participants. TMS was applied using a hand-held figure-eight coil with an outer winding diameter of 70 mm (MagstimCo., Whitland, Dyfed, UK). All pulses were delivered using a Magstim rapid stimulator connected to 4 booster modules that produce biphasic pulses. Stimulation was delivered over the motor cortical hand area with the coil placed tangentially to the scalp with the handle pointing posteriorly. The motor cortical hand area was defined as the location on the scalp where magnetic stimulation produced the largest motor-evoked potential (MEP) from the contralateral FDI when the participant was relaxed. The stimulation intensity was defined in relation to the active motor threshold (AMT) of the participant. The AMT was defined as the minimum intensity of single-pulse stimulation required to produce an MEP on more than 5 out of 10 trials from the contralateral FDI while the participant was maintaining a voluntary contraction of about 20% of maximum in the FDI.

The stimulation targets were localized using Brainsight, a computerized frameless stereotaxy system (Rogue Research, Montreal, Canada). This system uses an infrared camera to monitor the positions of reflective markers attached to the participant's head. Head locations are related in real time to the participant's previously acquired structural MRI data after the data are co-registered to a set of anatomical locations. Reflective markers are attached to the coil and the participant so that relative positions of the coil to the head (and the MRI) can be tracked, allowing precise positioning of the coil with respect to previously chosen cortical locations.

The repetitive TMS protocol used in the current study was identical to the continuous theta burst procedure described by Huang and colleagues (2005) and has been shown to depress activity in the stimulated region for up to 60 minutes following stimulation. Stimulation consisted of 50 Hz trains of three TMS pulses repeated every 200 ms continuously over a period of 30 seconds (450 pulses total).

Data Analysis

Participants’ accuracy rates were computed separately for each session for each metamemory category (remember, know, and guess). The effect of TMS on the task was evaluated by a repeated measures analysis of variance on accuracy with the stimulation site (vertex vs. frontal) and metamemory category as within-subject factors and frontal group (DLPFC vs VLPFC) as between-subject factors. Separate ANOVAs were run on the proportions of each metamemory response given by the participants. Post-hoc two-tailed t-tests were used to compare results between the two stimulation conditions where appropriate. Data for two participants were excluded because their memory accuracy did not exceed chance levels, so it could not be determined whether they were performing the
task or responding at random.
4.4 Results

Main experiment

There was a main effect of metamemory response ($F_{1,22} = 151.33$, $P < .001$) whereby participants were the most accurate overall when responding “remember” and were the least accurate when making a “guess”. Additionally, the results revealed a significant three-way TMS condition x metamemory x group interaction ($F_{1,22} = 4.791$, $P < .05$) indicating that the effect of frontal TMS differed at the different levels of metamemory between the groups. Post-hoc analyses revealed significant two-way interactions of TMS condition and memory in both the DLPFC group ($F_{1,10} = 7.786$, $p = .012$) and the VLPFC group ($F_{1,12} = 4.974$, $p = .046$). Accuracy was significantly greater for “Guess” responses after DLPFC stimulation than after vertex stimulation ($t_{11} = 2.29$, $p < .05$) suggesting that implicit memory was improved following interference to explicit processes in the DLPFC (Fig. 3a). Among those participants in the VLPFC group, the interaction was driven by slightly more accurate “know” responses after frontal stimulation as compared to vertex stimulation, yet less accurate “remember” and “guess” responses (Fig. 3b).

a) DLPFC TMS Accuracy

![DLPFC TMS Accuracy](chart.png)
Figure 4.3 Task accuracy for each metamemory response.  a) The accuracy of “guess” responses was significantly higher following DLPFC TMS than after control TMS.  b) IFG TMS produced a significant interaction in accuracy between “remember” and “know” responses—remember accuracy suffered relative to control TMS while know accuracy increased slightly.

It must be noted that there was no significant change in the proportion of each metamemory response across any of the TMS conditions (Fig. 4.4). This suggests that it was not the case that enhanced performance following DLPFC stimulation was due to a shift in response confidence, but rather that when explicit processing was not useful at retrieval (“remember” and “know” responses) the effectiveness of implicit memory processing (“guess” responses) was improved.
Figure 4.4 The proportion of each metamemory response given in each condition. Neither (a) DLPFC TMS nor (b) VLPFC TMS shifted the response criterion for metamory judgments. Changes in accuracy between control and PFC TMS cannot be explained by a change in the number of “remember”, “know”, or “guess” responses.
4.5 Discussion

Despite decades of research examining the distinctions between implicit and explicit memory systems, the interrelationship between the two remains unclear. By disrupting the function of focal parts of the PFC, a brain region implicated in explicit memory processes, we were able to enhance recognition performance when participants rely on implicit memory processes in a two-alternative forced-choice test. This is the first evidence, to our knowledge, that specifically interfering with the neural explicit memory processes at focal sites within the brain can increase the efficacy of implicit memory.

According to dual-process theories of memory, explicit episodic retrieval can be subdivided into two separate processes: recollection and familiarity. While recollection is characterized as a slow, intentional process involving rich contextual details, familiarity is thought to be fast, automatic, and unintentional retrieval lacking any contextual features. Several models suggest that familiarity processes might be intricately related to implicit memory and that recognition accuracy might reflect the contribution of increased fluency of the reprocessing of previously studied information (Mandler, 1980; Yonelinas, 2002; Jacoby and Whitehouse 1989; Johnston, et al., 1991; Verfaellie and Cernak, 1999). Our results support the contention that while implicit memory and familiarity processes supporting recognition might operate concurrently under certain conditions, they are dissociable mnemonic processes that rely on distinct neural architecture (Voss and Paller, 2009).

The recognition memory task used in this experiment was specifically designed to provoke responses based on implicit memory in the absence of explicit memory (Voss et al., 2008). At test, participants were presented with a pair of kaleidoscope images with a very high level of similarity across a variety of perceptual features and were forced to designate one of them as old. As a result, semantic and elaborative encoding and retrieval strategies, which are strongly associated with explicit memory (Yonelinas, 2002), could not be used effectively for this task. While recognition tests have usually been deployed to measure explicit memory, the metamemory responses collected in this task allow the examination of the contribution of implicit visual memory to recognition performance.

The fact that recognition responses assigned a “guess” metamemory response were more accurate than expected by chance might lead to the conclusion that these responses are not indicative of implicit memory per se, but rather simply low-confidence explicit familiarity judgments. In other words, the increase in accuracy of “guess” responses following DLPFC TMS is due to a shift in response confidence and not to an improvement in implicit memory performance. However, there was no increase in the proportion of “guess” responses made following DLPFC TMS arguing against this simple conclusion. Correct responses
that were previously assigned to the “know” metamemory category were not merely transferred into the “guess” metamemory category.

While we conclude that explicit memory processing in the frontal cortex can interfere with implicit memory, it is clear that DLPFC stimulation and VLPFC stimulation had at least somewhat different impacts on memory performance across different levels of metamemory. Unfortunately, our results cannot provide firm conclusions regarding the dissociation of these two regions with regard to their role in explicit memory. While it may be surprising that stronger effects were not found following VLPFC stimulation given the evidence of its role in both item memory and elaborative encoding (Blumenfeld and Ranganath, 2007; Badre and Wagner, 2007), our results do suggest that disrupting function in this region negatively impacts recollection accuracy and favors familiarity-based recognition. Additionally, because the effects of theta-burst stimulation likely lasted the duration of the experiment, we are not able to dissociate prefrontal contributions to encoding and retrieval processes. Further work could shed light on whether explicit memory processes that interfere with implicit memory occur primarily during study or at test.
Chapter 5. Inducing improvements in the Stroop task by reducing response inhibition via rTMS to the right inferior frontal gyrus

5.1 Overview

The right inferior frontal cortex has been implicated in the inhibition of task sets and responses. Recent neuroimaging work has shown that the right inferior frontal gyrus (rIFG) is the only region to show increased connectivity with word processing regions in the infero-temporal cortex known to show suppressed activity during color naming in a color-word Stroop task. This suggests that the rIFG could be a source of a top-down signal underlying this posterior modulation. In this study, 1-Hz repetitive transcranial magnetic stimulation (rTMS) was applied to disrupt the function of the rIFG in healthy participants prior to the performance of a modified color-word Stroop task. Participants were instructed to respond with a button press the printed color of words they saw on the screen. Each of four colors (blue, red, green, yellow) was mapped to a single button that participants learned before testing. Each run consisted of four blocks of incompatible trials (color and text mismatched, e.g. "blue" written in red) and four blocks of neutral trials (names of animals as the text).

Despite the fact that the Stroop task has classically been used to examine cognitive control function, following the TMS disruption of the rIFG, participants showed a decrease in overall response times across trial types without a significant difference in the number of errors produced. Additionally, rTMS seemed to have the greatest impact on the response times of compatible trials. These results suggest that participants engage in an unnecessary excess of response inhibition on the Stroop task and that disrupting frontal lobe function can lead to a paradoxical improvement in performance.
5.2 Introduction

The prefrontal cortex (PFC) has long been implicated in cognitive control functions that enable individuals to coordinate thoughts and actions in accordance with internal desires and goals. One of these “top-down” operations, often called selective attention, involves the filtering out of task-irrelevant information for the optimal processing of task-relevant information. Response inhibition is another form of cognitive control and is usually defined as the process by which a routine or pre-potent response is deliberately withheld or slowed.

Since its development decades ago, the Stroop Color Naming task (Stroop 1935; MacLeod 1991) has been used to illustrate fundamental cognitive control operations such as those described above. In this task, subjects are presented with a series of words and they must report the color in which these words are written. In order to perform this task, it is necessary to selectively attend to a single attribute of each stimulus (its color). This is especially true in critical trials in which the word that is displayed conflicts with the color in which it is written (e.g., the word BLUE displayed in red ink). The strong automatic tendency to read the word often competes with the naming of the color of the word and this “interference” must be overcome by selecting the weaker goal-relevant response in the face of competition from the stronger, yet task-irrelevant one.

Numerous neuropsychological studies have shown that patients with damage to different regions of the PFC show sizeable deficits in color naming in the Stroop task (Perret 1974; Vendrell et al., 1995; Stuss et al., 2001). Additionally, several other patient populations have been shown to have greater error interference effects relative to healthy adults (Abramczyk et al., 1983; Perlstein et al., 1998). More recently, neuroimaging studies have sought to specifically delineate which regions of the PFC contribute to successful performance on the Stroop task (Carter et al., 1995; Leung et al, 2000; Zysset et al., 2001) and to ascertain which cognitive control operations are carried out by which prefrontal regions and networks (Duncan and Owen 2000). Almost all of these neuroimaging studies have implicated the right inferior frontal gyrus (rIFG) as a region that is consistently activated when participants are asked to resolve the interference of an incongruent color-word pair.

Unfortunately, the specific operations carried out by the rIFG in cognitive control tasks remain unclear (Duncan and Owen 2000; Hampshire et al., 2009; Hampshire et al. 2010; Verbruggen et al. 2010). On the one hand, there has been mounting evidence that this region of the PFC is essential for response inhibition (for a review, see Aron et al., 2004). For example, studies employing stop-signal or go/no-go tasks that involve the cancellation of routine responses following an infrequent stop cue consistently find high levels of activity in the rIFG during stop trials (Logan and Cowan, 1984; Konishi et al., 1999; Rubia et al.,
2003, Aron et al., 2004). Similarly, studies looking at patients with focal brain damage seem to suggest that this region is uniquely important in the prefrontal cortex for these tasks (Aron et al., 2003; Hodgson 2007). In contrast, carefully controlled neuroimaging studies have suggested that the activity that is commonly observed in the rIFG during response inhibition tasks cannot specifically be attributed to generating inhibitory signals and might play a broader role in selective attention and target detection (Hampshire et al., 2010).

In the Stroop task, it seems clear that both selective attention and response inhibition are necessary components of successful task performance. While it is essential that subjects attempt to selectively attend to a single feature of the stimulus (i.e., its color and not the word meaning), the automaticity of word reading necessitates an inhibition of the response that coincides with the word meaning.

The aim of this work was to use repetitive transcranial magnetic stimulation (rTMS) to transiently disrupt function in the rIFG of healthy adult individuals prior to the performance of the Stroop task in order to elucidate the cognitive control operations executed in this region. If the rIFG is broadly important for selective attention, response times should be slowed and/or error rates should increase after rIFG disruption. However if this region is primarily aiding in response inhibition, response times should actually decrease during the Stroop task, as participants will be unable to inhibit automatic responding in the face of expected interference. In this latter case it must be noted that although error rates might increase due to overly quick responding, it is also possible that accuracy remains unchanged if subjects are usually unnecessarily slowing. This would lead to the paradoxical result in which disrupting a cognitive control mechanism actually improves task performance.
5.3 Materials and Methods

Participants
Twenty-two participants were recruited for this experiment. Four participants did not complete the rTMS session because of the aversive nature (muscle twitches, etc.) of rTMS over the rIFG. Additionally, one participant had response times greater than two standard deviations slower than the mean, so his data were excluded, resulting in a final sample of 17 participants. All participants had normal or corrected-to-normal vision, and none of the participants was colorblind (by self report). Additionally, all participants had a right hand preference and were native English speakers, and were paid for participation.

Experimental Paradigm

Participants performed a color word Stroop task (Stroop, 1935). Color words (blue, red, green, yellow) or animal names (horse, bird, cat, dog) printed in different colors (blue, red, green, yellow) were presented on the left side of the computer screen. Participants had to indicate the font color by pressing one of four buttons. For ease of task performance color-to-button mappings were presented at the bottom part of the screen throughout the duration of the experiment. Participants used four fingers of their right hand for responding with color-to-button mappings randomly assigned to participants. Compatible, neutral, and incompatible trials were presented with equal probability. In compatible trials, color and word were the same. In neutral trials, the task-irrelevant dimension (e.g., word meaning) was not related to the task (e.g., animal names). In incompatible trials, color and word differed. Each Stroop session was ten minutes long and was comprised of 8 blocks of 36 trials each. The stimuli were presented for 300 ms with an interstimulus interval of 1700 ms (Fig. 5.1).

Figure 5.1 Stroop task timing. Stimuli were presented for 300 ms with a 1700 ms intertrial interval. Congruent, incongruent, and neutral words were presented with equal frequency across each
block of the task.

After a training session to familiarize participants with the task and the color-to-button mappings, participants completed a baseline session of the task. Then, 1 Hz rTMS (600 pulses) was applied for ten minutes either over the rIFG (9 participants) or the vertex of the skull (8 participants). Immediately following stimulation, participants underwent another task session to assess how disruption affected task performance (Fig 5.2).

Figure 5.2 Experimental timeline. Each participant initially trained on the task for 10 minutes to learn color-to-button mappings and to familiarize themselves with the experiment. Following a baseline session, 1 Hz rTMS was applied to either the RIFG or the vertex of the skull. Immediately after stimulation, participants underwent another session of the task to assess performance changes due to rTMS.

Structural MRI scans were acquired in a previous session for all participants to aid in stereotaxic localization of rTMS.

rTMS Application
rTMS was applied using a vacuum-cooled figure 8 coil (5-cm diameter) powered by a Magstim Super-Rapid stimulator (Magstim Co., Whitland, South West Wales, UK). rTMS stimulus intensity was set at 80% of maximal output of the stimulator. The target sites were localized using Brainsight, a computerized frameless stereotaxy system (Rogue Research, Montreal, Canada). This system uses an infrared camera to monitor the positions of reflective markers attached to the participant's head. Head locations are related in real time to the participant's previously acquired structural MRI scan. Reflective markers are attached to the coil and the subject, so that relative positions of the coil to the head (and the MRI) can be tracked, allowing precise positioning of the coil with respect to previously chosen MRI locations. rTMS produced aversive side effects in some subjects. Due to the frontal stimulation site, repetitive contractions of face and jaw muscles occurred. Of the seventeen participants who completed the study, three reported mild to moderate headaches or scalp pain at the end of the session. No participants reported any impairment in concentration on the task.
due to discomfort. No other side effects were reported and there were no seizures.

rTMS Targets

The frontal stimulation site in the present study was chosen based on fMRI data recorded during performance of the Stroop task in our laboratories. In this study we found a negative relationship between activity in the rIFG and posterior sensory areas in the service of selective attention (unpublished data). We used a cluster from group level analysis (x=60, y=23, z=18) and constructed a spherical region-of-interest with a 5 mm radius in the native space of each participant’s structural MRI scan (Fig. 5.3).

![Figure 5.3](image_url) Frontal TMS target. Right IFG coordinates from a previous fMRI study were reverse normalized into each participant’s native-space structural MRI scan (representative subject is shown). Vertex rTMS locations were based on structural measurement of the skull and not MRI scans (not shown).

The rTMS target chosen as a control site was the vertex of the skull. This region was defined as the top of the skull on the midline between the nasion and the inion. Vertex sites are commonly used in the behavioral TMS literature as a control for any general, nonspecific effects of rTMS including general arousal, auditory click sounds, and the sensation on the scalp.

Data Analysis
The inclusion of a neutral condition allowed us to calculate three indices of
selectivity of information processing (attention). In most studies, the compatibility effect is calculated as incompatible minus compatible. Facilitation, i.e. speeded responses due to task-irrelevant compatible information, can be calculated as compatible minus neutral. Interference, i.e. slowed responses due to task-irrelevant incompatible information, can be calculated as incompatible minus neutral. We conducted separate repeated-measures ANOVAs using these three indices for error rates and median response times.

We considered the first task block in both pre- and post-TMS Stroop sessions as warm-up and excluded data from these blocks from the analysis. Error trials were also excluded from the response time analysis.
5.4 Results

Behavioral Results

There was a main effect of compatibility level on median response times (F\(_{1,15}\) = 23.94, p < 0.001), indicating that responses in incompatible trials were the slowest, followed by neutral trials, with compatible trials having the fastest RT. There was also an interaction between rTMS group and session (F\(_{1,15}\) = 4.87, p < 0.05). Post-hoc t-tests revealed that participants who received rIFG rTMS had significantly decreased RT relative to their baseline session (t\(_8\) = 2.34, p < 0.05), while participants who received vertex rTMS showed no difference in RT (t\(_7\) = 0.095, p = 0.93; see Fig. 5.4).

![Figure 5.4 TMS RT – Baseline RT (all trial types). Reaction times were significantly decreased across all trial types after right IFG rTMS. There was no change in RT after control TMS.](image)

Additionally, there was a trend toward a three-way rTMS group x session x compatibility interaction (F\(_{1,15}\) = 3.23, p < 0.1). This interaction was driven by the fact that participants who received rIFG rTMS seemed to have the largest decrease in RT in compatible trials suggesting that rIFG disruption led to the greater facilitation effects (Fig. 5.5). Post-hoc t-tests on the difference in RT between baseline and test Stroop sessions revealed that participants who received rIFG rTMS had a significantly greater decrease in RTs for compatible trials after stimulation than did participants receiving vertex rTMS (t\(_{15}\) = 2.18, p <
While participants receiving rIFG rTMS also had faster RTs in incompatible and neutral trials relative to their baseline session, this change in RT was not significantly different than that of vertex rTMS participants (incompatible trials - \( t_{15} = 1.72, p > 0.1 \); neutral trials - \( t_{15} = 1.52, p > 0.1 \)).

**Figure 5.5** TMS RT – Baseline RT broken down by trial type. While reaction times were numerically decreased when examining each trial type separately after right IFG rTMS, this effect was the greatest for compatible trials where a significant decrease was observed. There was no change in RT for any trial type after control TMS.

There were no significant effects on error rates in any condition. This suggests that the decrease in reaction times following rIFG rTMS did not lead to more errors and can therefore not be explained by a speed-accuracy tradeoff.
5.5 Discussion

Disruption of the rIFG prior to the performance of the Stroop task led to a paradoxical improvement on the task as evidenced by a decrease in response times without a concomitant increase in error rate. Furthermore, participants showed greater facilitation effects when faced with congruent stimuli with matching color-word pairs.

Other TMS studies targeting the rIFG during stop-signal and go/no-go tasks have frequently observed that disruption of this region leads to impaired ability in response inhibition tasks (Chambers et al., 2006; Chambers et al., 2007; Verbruggen et al. 2010). Similarly, patients with damage to the rIFG show deficits on these types of tasks (Aron et al., 2003). Usually, these results have been used to support the assertion that the rIFG plays a role in the actual inhibition of pre-potent responses. Our data seem to be in agreement with these reports. However, several researchers have suggested that the deficits seen on these tasks do not reflect a failure response inhibition per se, but rather an inability to detect salient cues that should affect current task plans (Hampshire et al., 2009; Hampshire et al., 2010; Greening et al., 2010; Verbruggen et al., 2010). These accounts suggest that the rIFG is important for updating action plans in accordance with newly acquired information, especially in contexts with competing response options.

There seems to be some support for this more general role of the rIFG in cognitive control tasks. For example, ample evidence from neuroimaging studies implicates the RIFG in attentional switching and rule switching (Cools et al., 2002; Derfuss et al., 2005; Hampshire and Owen, 2006; Greening et al., 2010). This work has consistently shown that when individuals need to switch to a new task or respond based on a new set of rules, the rIFG displays increased activity levels. However, a recent rTMS study targeting the rIFG did not find any evidence that disrupting this region led to deficits in task switching (Verbruggen et al., 2010). Moreover, it has been suggested that task switching inherently involves an aspect of response inhibition whereby the previous task or rule must be inhibited in favor of the newer, more appropriate one (Aron et al., 2004).

While evidence from our study cannot fully specify the role of the rIFG is in action updating, it does seem clear from our results that this region does in fact aid in response inhibition. Response times across all compatibility levels decreased following rIFG disruption. Because subjects were familiar with the task and were aware that there would be many trials with incongruent color-word pairs, they were likely globally slowing down (i.e., inhibiting) all responses in order to protect against errors from automatic word reading. This universal slowing ensures that participants were able to overcome reflexive responding based on this seemingly
irrelevant information. However, this information can actually aid responses in the approximately 1/3 of trials for which color-word pairs are congruent. rTMS to the rIFG seems to “release the brakes” and allowed subjects to take greater advantage of the facilitation effect of congruent stimuli in the Stroop task, leading to significantly faster responses to these items.

Similar effects have been found in various stop-signal tasks in which rTMS to rIFG has led to numerically or significantly faster response times on “go” trials (Chambers et al., 2006; Verbruggen et al. 2010). Additionally, it has been observed that schizophrenic patients also exhibit greater facilitation effects in the Stroop task than healthy adults, which has been explained by deficits in cognitive control mechanisms such as the inhibition of automatic responding and the selective attending of relevant information (Carter et al. 1992; Barch et al. 1999).

While response times were also decreased in incongruent and neutral trials following rIFG disruption, we did not observe any change in error rate. There are several possible explanations for this improvement in performance. Accuracy was quite high across all participants (>90% correct), which suggests that the variability in errors on the Stroop task may not be large enough to detect significant decrements in accuracy. In the same vein, participants may have misjudged the difficulty in the task and might have unnecessarily slowed their responses. During the initial training session, subjects became quite familiar with both the task and the button-to-color mappings. It is possible that they were capable of performing the task much more easily than they were aware of and that the conflict they experience during incongruent trials was relatively trivial. That is, responding based on the color has become the pre-potent response and does not require inhibition. If this were the case, then universally speeding all responses would not have a high cost and would lead to the superior performance we observed.

In conclusion, our results provide further evidence that the rIFG is essential for response inhibition. Additionally, producing deficits in inhibiting pre-potent responses via rTMS disruption to the rIFG leads to a paradoxical improvement on Stroop task performance. While our work cannot rule out that the rIFG aids in other cognitive control operations, the lack of any performance deficit following disruption to this region would seem to rule out its role in selective attention in this task.
Chapter 6. Concluding Remarks

The results presented in the preceding chapters have provided a number of insights into the mechanisms of top-down executive control. A summary of the central findings is discussed below.

Chapter 2 provided an overview of the existing literature that has sought to combine disparate neuroscience techniques to enrich our understanding of PFC function. While most theoretical models of prefrontal function have been based on indirect evidence that details how the PFC functionally interacts with distributed brain regions, this does not reflect a shortcoming on the part of those proposing the models. It is difficult to provide causal evidence of the role of the PFC using a single current cognitive neuroscience methodology. Nevertheless, I have reviewed the work of several groups that have used a combination of techniques to address the shortcomings inherent in each in order to provide a more nuanced understanding of the role of the PFC in executive function.

The work in Chapter 3 provided a direct example of how combining two such cognitive neuroscience methods, TMS and fMRI, can further our understanding of the mechanism by which the PFC performs top-down control operations. Continuous theta-burst TMS was shown to disrupt the efficacy of activity in the left inferior frontal cortex in aiding performance on a working memory task. This disruption also led to downstream changes in the selectivity of responses in extrastriate cortex. Moreover, this change in selectivity was tied to behavioral performance, providing a strong causal link between top-down control processes of the PFC and bottom-up visual processing in extrastriate cortex in the service of working memory. The collection of fMRI data during task performance also allowed an exploration of the whole brain, which resulted in evidence that provides clues as to how compensatory mechanisms in distant brain regions can overcome perturbations in focal brain activity. Taken together, Chapters 2 and 3 have shown how our understanding of prefrontal control can be greatly deepened by an amalgamation of available cognitive neuroscience tools.

Chapter 4 began an exploration of how prefrontal top-down control can hinder successful performance in circumscribed situations. While the PFC has long been known to be crucial for the ability to regulate thought and behavior in accordance with goals and task demands, its contribution to performance is not always positive. In this chapter, I examined a task in which attentive encoding has previously been shown to have deleterious impacts on visual recognition memory. By using TMS to disrupt two separate focal sites within the PFC prior to memory encoding, I was able to provide evidence that top-down control signals emanating from the frontal lobe interfere with superior memory performance. Unfortunately, strong dissociations between frontal regions could not be uncovered. While it is clear that the disruption of dorsolateral PFC led to different behavioral outcomes than the disruption of ventrolateral PFC, a clear delineation...
of the processes that each region contributes to task performance was impossible to ascertain. Further work using this task combining both TMS and fMRI could aid in elucidating these underlying top-down control mechanisms.

Another task in which the disruption of prefrontal cortical function led to improvements in behavioral performance was described in Chapter 5. The Stroop task has been used for decades as a measure of cognitive control. When faced with incongruent stimuli, it is necessary to suppress the response based on the automatic reading of the words in order to correctly respond based on the ink color of each stimulus. However, the work described here presents evidence that there might be an excess of top-down control exerted by the PFC on this task and that disrupting activity in the right IFG can lead to behavioral improvement. These results were discussed in light of the literature that implicated the right IFG in response inhibition. An excess of response inhibition seems to lead to an unnecessary slowing of response times and by releasing this inhibition, these response times decreased without an increase in error rate. Furthermore, lessening response inhibition allowed participants to take greater advantage of facilitation effects of compatible stimuli, in which a normally irrelevant dimension can aid task performance.

By describing conditions in which perturbing function in the PFC can both help and hinder task performance the original work in this dissertation has sought to reveal a more complete understanding of the underlying mechanisms of top-down executive control.
References


Luria, A.R. (Eds.), Psychophysiology of the frontal lobes (pp. 3-26). New York: Academic Press.


