Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Signature:

_____	_____
Jared L. Moreines	April 20th, 2010

Dissociating the Effect of Treatment-Resistance on Neuropsychological Performance in Patients with Major Depressive Disorder

by

Jared L. Moreines

Adviser

Paul E. Holtzheimer, M.D.

Department of Neuroscience and Behavioral Biology

_____
Paul E. Holtzheimer, M.D.
Adviser

_____
Helen S. Mayberg, M.D.
Committee Member

_____
Amy Alderson, Ph.D., ABPP-CN
Committee Member

_____
Keith Easterling, Ph.D.
Committee Member

_____
Melvin Konner, M.D., Ph.D.
Committee Member

_____
April 14th, 2010

Dissociating the Effect of Treatment-Resistance on Neuropsychological Performance in Patients with Major Depressive Disorder

By

Jared L. Moreines

Adviser
Paul E. Holtzheimer, M.D.

An abstract of
A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Neuroscience and Behavioral Biology

2010

Abstract

Dissociating the Effect of Treatment-Resistance on Neuropsychological Performance in Patients with Major Depressive Disorder

By Jared L. Moreines

Deficits in cognitive functioning are common in depression and may vary according to a patient's course of illness. Differences in resting functional magnetic resonance imaging (rfMRI) Functional Connectivity (FC) of brain regions involved in mood and cognition have been shown to be associated with cognitive ability and depressive state. In the present study, healthy controls and individuals with depression who were either treatment-naïve or severely treatment-resistant underwent rfMRI scanning and neuropsychological testing on subtests from the Cambridge Neuropsychological Test Automated Battery and standard and emotional versions of the Stroop Task, all previously shown to be deficient in individuals with depression. Comparisons between all patients and controls revealed differences in measures of processing speed, bias for negative information, and executive functioning. Comparisons between the treatment-naïve and treatment-resistant groups revealed differences in processing speed of directed response to emotional words. Due to an issue in the rfMRI data, it could not be analyzed for this thesis. However, the results of neuropsychological testing suggest select differences in the neurocognitive profiles of treatment-naïve versus treatment-resistant depression related to processing speed. Limitations of the current study include inability to dissociate past and current medication effects from treatment-resistance, small sample size, and no correction for multiple comparisons.

Dissociating the Effect of Treatment-Resistance on Neuropsychological Performance in Patients with Major Depressive Disorder

By

Jared L. Moreines

Adviser
Paul E. Holtzheimer, M.D.

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Neuroscience and Behavioral Biology

2010

TABLE OF CONTENTS

BACKGROUND

Major Depression and Treatment Resistance

Major depression is a psychiatric disorder associated with pervasive and persistent feelings of sadness or loss of interest in previously pleasurable activities, lasting at least two weeks in duration. Additional symptoms can include decreased energy, suicidal ideation, impaired concentration, changes in eating and sleeping patterns, and decreased libido (American Psychiatric Association, 2000). A large national survey (National Comorbidity Survey Replication) with a large, community-based participant group highly representative of the U.S. population (N=9,282) estimated the 12-month prevalence of Major Depressive Disorder (MDD) in the U.S. to be 6.7%, and the lifetime prevalence to be 16.6%, with the rate for women roughly twice the rate for men (Kessler, Berglund, et al., 2005; Kessler, Chiu, Demler, Merikangas, & Walters, 2005). The initial onset of MDD in respondents generally occurred in early adulthood (median age = 32, mid-50th percentile = 20-40), but ranged from adolescence to old age. Highlighting the large worldwide impact of MDD, the World Health Organization ranked unipolar major depression as the third leading contributor to overall burden of disease worldwide in 2004, and predicted the rank to increase to number one by the year 2030 (World Health Organization, 2008).

Despite a large number of medications currently available, MDD has a relatively high rate of treatment failure (Nemeroff, 2007). Based on data from STAR*D (a large community-based treatment study of individuals with MDD), up to two thirds of patients do not achieve remission after an initial course of antidepressant treatment (Trivedi, Rush, et al., 2006), and one third do not achieve remission even after a sequence of up to

four consecutive treatments (including either a switch to a new medication or augmentation with a second medication) (Rush, et al., 2006; Trivedi, Fava, et al., 2006). Although new, well-tolerated, neuromodulation therapies (e.g. Transcranial Magnetic Stimulation, Vagus Nerve Stimulation, Deep Brain Simulation) are being developed for individuals with treatment resistant depression (TRD) (Moreines, McClintock, & Holtzheimer, In Press), more research is needed to understand the pathophysiology and symptomatology specific to TRD given its high prevalence and increased patient suffering, disability, treatment costs, and suicide risk (Crown, et al., 2002; Nemeroff, 2007).

Neuropsychological Functioning in Depression

While depression is most notably associated with reductions in mood, deficits in neurocognitive functions are also common. These can include impairment in a variety of traditional neuropsychological domains, including divided attention, short-term and long-term memory, and executive functions (e.g. cognitive flexibility, inhibition, and problem solving) (McClintock, Hussain, Greer, & Cullum, 2010). In addition, MDD can be associated with specific deficits and patterns of cognitive performance unique to individuals with MDD. These include hypersensitivity to negative feedback (i.e. increased likelihood that one error during task performance will be followed by another) (Elliott, et al., 1996), and a bias towards material with a negative emotional valence (exhibited by faster response to negative versus positive cues during emotional face or word processing tasks, and greater likelihood of remembering negative portions of a verbal recall test) (Erickson, et al., 2005; Murphy, et al., 1999). However these deficits

are not present in all individuals, and results across studies have been inconsistent in the specific pattern of deficits identified, with some studies finding severe and widespread deficits, and others failing to identify any.

In an early attempt to characterize the neurocognitive profile of depression and explain the large variation in the reported findings, a meta-analysis by Veiel (1997) proposed a "global-diffuse" impairment across many cognitive domains. In this model, simple tasks such as simple reaction time and attention showed infrequent and mild impairment, and more complex tasks such as executive functioning showed more common and severe impairment. Although the essence of this theory has retained its validity, subsequent studies have attempted to clarify more definitively the sources of the variability observed in patients' neuropsychological performance.

Recently, many studies have investigated the impact of depressive illness severity as a determinant of neuropsychological performance (McClintock, Husain, Greer, & Cullum, 2010). Subsequently, a large number of illness characteristics that may affect the level of cognitive impairment that a patient experiences have been identified. For example, cognitive deficits may be more severe in patients who have suffered a greater number of previous episodes (Paelecke-Habermann, Pohl, & Leplow, 2005) and in patients who have experienced a higher number of hospitalizations related to their illness (Harvey, et al., 2004; Purcell, Maruff, Kyrios, & Pantelis, 1997). In a recent systematic review of neurocognitive data in individuals with depression published between 1980 and 2008, McClintock et al. (2010) concluded that the majority of available data suggest a strong association between greater illness severity and more profound cognitive deficits, though it remains unclear which specific measures of severity (e.g. score on standardized

clinical rating scales, duration of past and current episodes, number of lifetime depressive episodes, mood-related hospitalizations, suicide attempts, and treatments attempted) are the most important determinants of neuropsychological impairment.

In a recent attempt to provide a quantitative role of depressive illness severity in determining an individual's neuropsychological performance, Gorwood et al (2008) tested a large (N=8,229) sample of depressed outpatients on delayed paragraph recall before and after treatment and used structural equation modeling to quantify the relative risk associated with different illness features. They determined that 1) at the pre-treatment test date, higher scores on depression and anxiety rating scales proved the greatest predictors of memory performance, 2) at the follow-up test date (when many patients had experienced some level of recovery from their illness), number and length of prior depressive episodes better predicted memory performance than did current depression and anxiety rating scale scores, and 3) each prior depressive episode (up to the fourth) predicted a 3% decrease in memory performance (Gorwood, Corruble, Falissard, & Goodwin, 2008). This suggests that with adequate power, modeling of the differential impacts of various components of illness severity may be possible for other neuropsychological test domains. However, delayed paragraph recall is a well-established test of hippocampal functioning (Buckner, et al., 1995), and hippocampal atrophy is highly unique as being a structural change associated with depression with a direct correlation with duration of illness (Sheline, Wang, Gado, Csernansky, & Vannier, 1996). Thus, due to a lack of other structural (or functional) illness markers with both established neuropsychological test correlates and reliable associations with illness severity, it is unlikely that a similar level of modeling is possible for other

neuropsychological domains. However, this study provides important quantitative support for the relationship between depressive illness severity and neuropsychological performance.

Taken in sum, the results of the cumulative literature regarding illness severity suggest that individuals with tendency for recurrence, or those who have been depressed for a greater amount of time due to increased severity or treatment resistance, may demonstrate greater cognitive deficits than those with a shorter or more limited course of illness. The direct comparison remains to be published, however, between depressed individuals naïve to treatment and depressed individuals who have failed multiple treatment attempts, and are thus classified with severe treatment resistance. Such a comparison could suggest whether one's stage of illness (treatment-naïve or treatment-resistant) affects their neuropsychological profile.

Functional Neuroimaging in Depression and Neuropsychology

While patterns of deficits in neurocognitive performance alone can suggest potentially abnormal functioning of isolated brain regions, the observed relationship between neurocognitive functioning and depressive illness severity implies dysfunctional communications between the brain regions involved in the performance of neurocognitive tasks and those implicated in depression. To evaluate this hypothesis most effectively, one must explore functional interactions between these regions in individuals with demonstrated deficits.

The pathophysiology of depression has been studied using a number of structural, resting-state functional, and task-based functional neuroimaging techniques. Some of the

abnormalities identified include structural changes such as decreased hippocampal volume and increased amygdalar volume, resting-state functional changes such as hypoperfusion of the prefrontal cortex (with possible hyperperfusion of select prefrontal regions such as the subgenual cingulate), and task-related functional changes such as enhanced amygdalar response to fearful faces (Mayberg, 2009).

Positron Emission Tomography

One commonly used functional imaging modality for depression research has been resting state positron emission tomography (PET) – an imaging technique that utilizes radioactively labeled water or glucose to characterize regional blood flow or metabolism, respectively. This technique has been particularly helpful over the years to study the relative function of specific brain regions in patients with depression (compared to non-depressed controls) because it allows for the identification of regions with abnormal activity even when individuals are at rest. These studies have identified a consistent set of brain regions that appear to be involved in the neurobiology of depression and antidepressant treatment response, leading to the development of a novel treatment for depression: deep brain stimulation (DBS) of the subcallosal cingulate white matter (SCCwm) for severe TRD (Mayberg, 2006).

PET imaging has also been used to specifically study the neurocognitive deficits in individuals with depression (as well as in other disorders). These studies have employed both resting and task-based paradigms. Goldstein et al (2004) used PET imaging to measure resting metabolism in individuals with addiction to alcohol or cocaine compared to healthy controls. They found that reduced prefrontal cortex glucose

metabolism in the cocaine addiction group during rest was associated with neurocognitive deficits in this group in attention and executive functioning (Goldstein, et al., 2004). Similarly, Dao-Castellana et al (1998) found that alcoholics' metabolism in the mediofrontal cortex at rest correlated with their performance on a test of verbal fluency, and their metabolism in the dorsolateral prefrontal cortex at rest correlated with their performance on a task of executive functioning and response inhibition (Dao-Castellana, et al., 1998). In a task-based paradigm, Elliot et al (1997) used PET imaging to study activation patterns in individuals with depression as they completed a test of executive functioning and found that depressed individuals showed an overall attenuation in task-related neuronal function that was most prominent in the caudate, thalamus, anterior cingulate, dorsolateral prefrontal cortex, and ventrolateral prefrontal cortex (Elliott, et al., 1997).

A limitation of these studies is that PET imaging has low temporal resolution: i.e., activity is averaged over a few minutes (for blood flow) to several hours (for metabolism). Analyses comparing the "functional connectivity" of brain regions are therefore limited to group-level analyses rather than patient-level analyses. As depression has increasingly been recognized as the dysfunctional interaction of a number of brain regions acting in concert rather than dysfunction of any one region in isolation, methods that can better target such functional interactions are becoming more necessary.

Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) has much better temporal resolution compared to PET and has been utilized for task-activated studies of

neuropsychological performance for many years. Abnormal results from these studies can be classified into two groups: those that showed normal task performance in the presence of task-related over-activation (Fitzgerald, et al., 2008), suggesting a pattern of "cortical inefficiency" (Harvey, et al., 2005), and those that showed impaired performance in conjunction with decreased activation in a set of regions (Hugdahl, et al., 2004). This consistent bifurcation suggests that impaired task performance in individuals with depression is reflected in reductions of cortical activation, and normal performance requires a greater level of cognitive activation (Thomas & Elliott, 2009).

One undesirable aspect of task-based fMRI studies is the level of variability in activation patterns associated with minor adjustments in task design, even when two tasks are targeting identical neurocognitive domains. Experiments are therefore extremely vulnerable to the idiosyncrasies of the specific task used; this means that different tasks apparently measuring the same cognitive domain may yield different results (this effect has been described in depth as it pertains to studies of the dorsolateral prefrontal cortex (Fitzgerald, et al., 2006)). Until recently, studies using fMRI were limited to these such task-based activation paradigms, leaving PET imaging as the only source for high spatial-resolution imaging of group differences in resting brain function.

Resting State Functional Magnetic Resonance Imaging

A significant advance in fMRI-based research came when Biswal et al (1995) observed that resting-state spontaneous activations in hand sensorimotor cortex (that were previously thought to be noise or artifact) were correlated with spontaneous activations in other brain regions related to movement (Biswal, Yetkin, Haughton, & Hyde, 1995).

Since then, the burgeoning field of resting-state fMRI (rfMRI) Functional Connectivity (FC) research has emerged, exploring networks of functional connections within the brain at rest (Fox & Raichle, 2007), something previously only possible using PET imaging. In order to conduct an rfMRI FC analysis, an experimenter defines a seed Region of Interest (ROI) based on an a priori hypothesis about that region's FC network that is to be tested by that study. The time course of this ROI's spontaneous activations during the rfMRI scan is then correlated with those for every other voxel in the brain to determine which regions had similarly timed spontaneous fluctuations, suggesting the two regions may be "functionally connected." These rfMRI FC analyses allow one to compare neural network function– at rest –between different groups (e.g., patients vs. controls, pre- vs. post-treatment). For example, Greicius et al. (2008) used rfMRI FC to determine that individuals with depression may have greater contributions of the subgenual cingulate, a crucial node in the pathophysiology of depression, to the default mode network (DMN), a set of brain regions thought to be more active when a subject is at rest and less active when the subject is performing any type of goal-directed task (Greicius, 2008). More recently, another group used a more advanced modeling technique to incorporate the FC within a number of brain regions previously implicated in depression and were able to predict, with strong accuracy, whether an individual was healthy or depressed based on the FC relationships between these regions (Craddock, Holtzheimer, Hu, & Mayberg, 2009).

Numerous task-based fMRI FC studies (i.e. degree of co-activation of brain regions *during* task performance) have examined the neuronal networks implicated in cognitive performance deficits in individuals with depression. These have often

suggested a possible over-recruitment of certain functional connections in depressed subjects versus healthy individuals. For example, such studies have employed tasks of verbal working memory (Vasic, Walter, Sambataro, & Wolf, 2008), executive functioning, and response inhibition (Schlosser, et al., 2008), among others.

Just as studies using resting PET have explored the relationship between the resting brain and neuropsychological performance, the same can be done using rfMRI FC. In a study of healthy controls, Song et al. (2008) correlated subjects' whole-brain rfMRI FC of the dorsolateral prefrontal cortex (DLPFC, a region implicated in many executive tasks) with their performance on a verbal IQ task completed outside of the scanner. The authors concluded that in individuals with higher IQ, resting DLPFC activity was more synchronized with posterior brain regions (Song, et al., 2008). However, such rfMRI FC analyses have not yet been used to investigate rfMRI patterns associated with cognitive performance deficits in individuals with depression.

Present Study Intentions

In the present study, we evaluated neurocognitive functioning in healthy adults and adults with depression who were either naïve to treatment or chronically treatment-resistant. We then attempted to identify potentially aberrant neural connections associated with observed impairments in task performance. Subjects were tested on subtests from a standardized neurocognitive battery that were selected based on previous data suggesting that individuals with depression would show impairment in task performance. After identifying potential neurocognitive impairments in the patient groups, seeds for whole-brain rfMRI FC analyses were selected based on previous imaging literature identifying

regions most likely to be mediating the observed impairments in task performance. Based on a large body of literature regarding specific neuropsychological deficits associated with depression, we predicted depressed individuals would show some level of impairment (as compared to healthy controls) on all of the included tasks, with the chronically treatment-resistant group performing more poorly than the treatment-naïve group. Next, based on a growing collection of studies exploring the representation of cognitive ability in resting brain functional connectivity, we expected to find group differences in FC maps of seed ROI's selected for the between group comparisons. We predicted that poorer task performance in the depressed groups would be associated with reduced functional connections to other regions previously shown to be important for either task-specific performance or overall normal neuropsychological functioning. We also expected that these regions would show enhanced connections to brain regions implicated in the pathophysiology of major depression.

METHODS

Subjects

This study was approved by the Emory Institutional Review Board (IRB). All subjects provided written informed consent. Subjects were recruited from ongoing studies within the Emory University School of Medicine Department of Psychiatry and Behavioral Sciences between September 2007 and January 2010. Subject groups included Healthy Controls (HC) (no history of depression or other major psychiatric illness), treatment naïve depressed (TND) patients, and severely treatment resistant depressed (TRD) patients (including patients with treatment-resistant bipolar II depression). Structured diagnostic interviews and paper-pencil tests and rating scales (detailed below) were used to establish patients' current psychiatric status. For patients, consultation with study psychiatrists confirmed diagnosis. For all groups, subjects with major medical or psychiatric comorbidities were excluded. Demographic data (age, years of education completed, and IQ measured by the North American Adult Reading Test (NAART)) were collected. For the depressed groups, clinical data regarding their age of illness onset, duration of current depressive episode, number of prior depressive episodes (as well as hypomanic episodes for TRD patients with bipolar II disorder), past suicide attempts, and lifetime mood related hospitalizations were collected in order to control for confounding measures of illness severity between groups.

Healthy Controls

The HC group consisted of 15 healthy volunteers (9 males, 6 females; age 26-63, mean (SD) = 36.0 (10.7); education mean (SD) = 17.3 (2.2); IQ mean (SD) = 112.1 (6.1);

BDI mean (SD) = 0.80 (1.4)) recruited from ongoing studies comparing rfMRI and PET imaging between individuals with depression and healthy controls (R01MH073719 NIMH, PI: Helen Mayberg, MD and a NARSAD Young Investigator award to Paul Holtzheimer, M.D.). Subjects in these studies were recruited from the Atlanta community via advertisements and referral sources. These individuals met with the research staff to discuss their present condition, past medical, and psychiatric history to determine eligibility. Additional paper and pencil tests were administered to confirm that no depression or other medical, neurological, or psychiatric conditions were present. These included the Structured Clinical Interview for DSM-IV diagnoses (SCID-I) (First, Spitzer, Gibbon, & Williams, 2002) to systematically evaluate the presence of a psychiatric illness, and the Mini-Mental State Exam (MMSE) (Folstein, Folstein, & McHugh, 1975) to screen for the presence of dementia or globally impaired cognitive functioning. All healthy controls received a resting state fMRI scan as part of their participation in R01MH073719 or Dr. Holtzheimer's NARSAD.

Following completion of rfMRI scanning and confirmation of successful imaging data acquisition, subjects were then contacted based on previously expressed potential interest in participation in optional neuropsychological testing. Neuropsychological testing was arranged as close as possible to the date of rfMRI scanning for optimal consistency in patients' mood and cognitive state, but due to scheduling feasibility, time between rfMRI scanning and neuropsychological testing varied from a few hours to four months. To account for this, on the day of testing, participants completed the Beck Depression Inventory – II (BDI-II), a self-report questionnaire consisting of items

regarding the subject's mood, cognitive attitudes, sleeping, and eating patterns over the last two weeks, to confirm lack of depressive symptoms.

Treatment Naïve Depressed Group

The treatment-naïve group consisted of 12 individuals with treatment-naïve depression (TND) (7 males, 5 females; age 24-62, mean (SD) = 37.4 (12.3); education mean (SD) = 14.75 (2.0); IQ mean (SD) = 108.5 (9.7); BDI mean (SD) = 27 (8.2)). These individuals were recruited from The Emory CIDAR (P50 MH077083-01 NIMH, PI Helen Mayberg, MD), a large study within Emory's Department of Psychiatry and Behavioral Sciences examining biomarkers predicting antidepressant response in individuals with depression who have never previously sought treatment. Participants in this study were recruited from the Atlanta community via advertisements and referral sources. Inclusion criteria consisted of a current diagnosis of Major Depressive Episode (MDE) (as specified in the Diagnostic and Statistical Manual 4th edition text revision (DSM-IV-TR) (American Psychiatric Association, 2000)) determined by a structured diagnostic interview (SCID-I) with MDD being the primary diagnosis and a minimum score of 15 on the 17-item Hamilton Depression Rating Scale (HDRS-17) (Hamilton, 1960). Participants must not have received an adequate course of treatment (neither medication nor psychotherapy) for major depression at any point previously, even if they have had prior episodes of depression. Patients in this group underwent a resting state fMRI scan (before beginning treatment) as part of their participation in The Emory CIDAR.

Following completion of rfMRI scanning, subjects were contacted based on previously expressed interest in participation in optional neuropsychological testing. Neuropsychological testing was performed within one week of rfMRI scanning, before patients began treatment (one patient had received one session of cognitive behavioral therapy (CBT) and another had received 4 days of antidepressant medication, neither of which were deemed significant enough to warrant removal of these subjects). On the day of testing, participants completed the BDI-II to gauge current depression level.

Treatment Resistant Depressed Group

The treatment resistant depressed group consisted of 17 patients with severe TRD (7 males, 10 females; age 26-58, mean (SD) = 41.9 (9.0); education mean (SD) = 16.35 (2.9); IQ mean (SD) = 113.4 (5.9); BDI mean (SD) = 37.7 (11.3)). These patients were drawn from an ongoing study exploring the safety, efficacy, and potential mechanism of DBS of the SCCwm as a treatment for TRD. Participants in this study were recruited from the international community based on referral sources. Patients in this group must have a current MDE of at least two years in duration, diagnosed by SCID-I structured interview and confirmed by two independent psychiatrists, a minimum score of 20 on the HDRS-17, a maximum Global Assessment of Functioning of 50, and treatment-resistant depression, defined as failure to respond to a minimum of four different antidepressant treatments, including medications or evidence-based psychotherapy administered at adequate doses and duration during the current episode, and failure or intolerance of an adequate course of electroconvulsive therapy (ECT) during any episode (confirmed by medical records) or refusal of ECT due to a reason considered to be valid by the study

psychiatrist. Participants received a resting state fMRI scan (before undergoing DBS implantation surgery) as part of their participation in the DBS study. Neuropsychological testing was also performed before patients underwent DBS implantation surgery. Because these patients were meeting regularly with study staff during this baseline period, the BDI-II score recorded for participants from this group was the most recent completion of this rating scale for the DBS study (this was always within the 3 days prior to or following neuropsychological testing).

Neuropsychological Testing Procedure

Neuropsychological testing occurred in the Woodruff Memorial Research Building of Emory University, in a quiet room with slightly dimmed lighting to reduce interference from noise distraction and computer-screen glare. The CANTAB test battery was administered on a Dell Latitude D520 laptop computer (Dell, Inc; Round Rock, TX, USA) running CANTAB Eclipse (Cambridge Cognition Ltd; Bottisham, Cambridge, UK), the battery's standard administration software. Subjects responded to task cues using either the CANTAB's included press pad or an accessory Magic Touch transparent touch-screen (Keytec, Inc; Garland, TX, USA) placed over the screen of the laptop.

Testing was administered by trained research personnel, although this individual was not blinded to a subject's status as patient or control. Prior to each subtest, the experimenter read the test's standard instructions aloud to the participant and ensured the participant understood all instructions. Due to the use of a consistent order for subtest administration during completion of the neuropsychological battery by a large portion of the TRD patients, order of subtest administration was not counterbalanced, so as to

maintain consistency across groups. In the DBS study for which these TRD subjects were tested, the decision not to counterbalance was based on that study's primary use of the neuropsychological testing data to compare (with-in subjects) changes over time. Following completion of the computerized battery, subjects completed the traditional and emotional versions of the Stroop task. Finally, subjects completed the NAART, for which responses were recorded using a Sony ICD-P320 Digital Voice Recorder (Sony Corporation; Tokyo, Japan) for scoring at a later date. The entire neuropsychological testing procedure took approximately 2.5 hours. At the completion of testing, subjects were compensated $50 for their time (with the exclusion of individuals from the TRD group, for which testing was a component of their participation in a clinical trial of DBS of the SCCwm for severe TRD).

Neurocognitive Battery

The neuropsychological battery included select subtests from the computer-based neurocognitive test battery Cambridge Neuropsychological Test Automated Battery (CANTAB) (Cambridge Cognition Ltd; Bottisham, Cambridge, UK) as well as traditional and emotional versions of the Stroop task and the NAART. The CANTAB is a widely used and well-validated computer-based standardized neurocognitive battery (Robbins, et al., 1998). It is desirable for use in clinical trials with multiple research personnel because its computer-based design affords high inter-rater reliability and does not require the experimenter to be a certified psychometrist or clinical neuropsychologist.

The battery used in the present study was identical to that currently used in an ongoing study of SCCwm DBS for TRD, from which a portion of the patients in this

study were drawn. The set of subtests included were selected based on prior studies suggesting the specific test or the domain of neurocognitive functioning it targets may be impaired in individuals with depression (See Table 1. and text below for description of neurocognitive battery).

North American Adult Reading Test

The NAART is a measure of pre-morbid verbal IQ (Blair & Spreen, 1989). Its use is based on evidence that verbal IQ accounts for 90% of variance in intelligence, and that verbal IQ does not change with onset of depression. Thus, although it is administered *while* an individual may be depressed, the NAART provides a reasonable estimate of one's level of intellectual functioning *prior* to the onset of depression. This measure is included in order to evaluate for the presence of pre-morbid differences in intellectual ability that could account for any observed differences in neuropsychological performance. The other measure used to evaluate cohort differences in intellectual ability is subjects' total number of years of education completed.

Stroop Color-Word Task (Standard and Emotional Versions)

The Stroop Color Word task (Stroop, 1935) primarily assesses response inhibition and executive function. Subjects read down a series of lists of 100 words. Subjects are instructed to read aloud as many of the words possible in the time given to them (45 seconds for each trial). In the first trial, the words "red," "green," and "blue" are printed in black ink, and the subject is asked to simply read aloud through as many words on the list as possible in the allotted time. Subjects' speed on this control run can be used as a

proxy to measure their verbal processing speed. In the second trial, the characters "XXXX" are printed in either red, green, or blue ink, and the subject is asked to name the color of as many of the characters as possible in the allotted time. In the third trial, known as the interference trial, the words "red," "green," and "blue" are printed in either red, green, or blue ink (with no word printed in the color it describes), and subjects are asked to name the color of the ink of as many of the words possible in the allotted time, while ignoring the actual words printed on the sheet. For each trial, the number of words or colors (depending on the trial number) the subject names is recorded. Stroop interference effect is measured by the number of fewer words the subject was able to name in the incongruent color-word list compared to the color-only trial. In a prior study that transformed each sub-score to a measure of time per item named (45,000ms / sub-score total), individuals with depression were found to spend comparable time per word named (MDD=421.1, HC= 402.1 (p=0.344)), but greater time per color named (MDD=602.0, HC=508.0 (p=0.008)) and color-word combination named (MDD=1108.8, HC=867.0 (p=0.018)) as compared to controls, likely due to deficits in executive functioning and response inhibition (Harvey, et al., 2004), although this study did not report interference effect controlled for color naming speed.

For the Emotional Stroop, subjects are presented with one list each of positive, negative, and neutral words printed in red, green, or blue ink (with each word list containing words only of that emotional valence but printed in each of the different colors). Subjects are instructed to name the color of the ink the words are printed in, ignoring the actual words printed. It has been suggested that depressed subjects are influenced differently by the negative words than the positive or neutral words, as

demonstrated by the relative number of words they can name in the allotted time for each list. Individuals with depression have been found to complete less of the negative word list compared to the neutral word list, theoretically because they are more distracted by the negative words, and thus delay their color naming response (Segal, Gemar, Truchon, Guirguis, & Horowitz, 1995).

Motor Screening Test (MOT)

The Motor Screening Test (MOT) was used to 1) introduce subjects to the touch screen computer testing setting and 2) screen subjects for ability to follow commands, respond to visual cues, and produce directed movements. Subjects were asked to touch a series of cross hairs appearing sequentially in different locations on the screen. A short ascending series of tones informed the subject after each correct response. No tone was played if the subject missed the target. Results from this task were not analyzed either between groups or in conjunction with rfMRI data. Rather, if a subject could not complete the task, the remainder of the test battery was not attempted.

Affective Go/No-Go

The Affective Go/No-Go (AGN) tests for an affective/emotional bias (either positive or negative) in the processing of verbal information. Because this test includes dissociable elements of decision-making, response inhibition, set shifting, and overall verbal processing speed, affective bias in each of these domains can be calculated. The subject is shown a series of words appearing one at a time on the computer screen for 300ms with a 900ms inter-stimulus interval. Half of the words are positive (happy) and

half are negative (sad). The subject is informed before each round whether the "target valence" for that round is positive words or negative words (with the other being the "distracter valence"). Subjects are instructed to press the response pad as quickly as possible when the word currently appearing on the screen is of the target emotional valence, and to ignore words of the distracter valence. The test alternates with blocks of two rounds of each target valence, for a total of 10 rounds (5 blocks). The order of target valence (H=Happy, S=Sad) is HHSSHHSSHH. The first 2 rounds are for practice, leaving a total of 8 rounds to be analyzed, 4 of which are "shift" conditions (where the target valence is different from the prior round) and 4 of which are "non-shift" conditions (where the target valence is the same as in the prior round). Each target valence thus has two rounds of shift and two rounds of non-shift. Output measures (given by valence and calculated for shift conditions only, non-shift conditions only, and total across all conditions) include mean latency for correct responses, number of commission errors, and number of omission errors.

This test (or in some cases, a close variant) has been used in previous studies of individuals with depression (Elliott, Rubinsztein, Sahakian, & Dolan, 2002; Erickson, et al., 2005; Murphy, et al., 1999), and results have found individuals with depression to respond faster to negative words than positive words (mean H = 570ms, mean S = 557ms), whereas healthy controls (mean H = 542ms, mean S = 545ms) and individuals with mania (mean H = 554ms, mean S = 578ms) have both been found to respond to positive words more quickly than negative words (MDD vs. HC latency for positive words $F(1,65)=4.05$, $p<0.05$; latency for negative words p=N.S.; Manic vs. HC latency

for positive words p=N.S.; latency for negative words ($F(1,65)=4.87$, $P<0.05$)) (Murphy, et al., 1999).

Cambridge Gambling Task

The Cambridge Gambling Task (CGT) tests various aspects of risk taking. Subjects are shown a row of 10 boxes across the top of the screen, some of which are red, and some of which are blue, and told that a yellow token is hidden under one of the boxes. The subject is instructed to select whether they think the token is hidden under a red or blue box, and then bet a portion of their current point total based on the confidence of their decision. If they are correct, then their bet amount is added to their current point total, and if they are incorrect, then their bet amount is subtracted from their current point total. Subjects are instructed to build up as many points as possible. The subject begins each round with 100 points, and each round contains 10 trials. A total of eight rounds are completed. For the first four rounds, the initial bet presented to the subject is small, and they must wait for the number to progressively increase if they want to bet a large amount. For the second four rounds, the initial bet presented to the subject is large, and they must wait for the number to progressively decrease if they want to bet a small amount. Because of this design, it is possible to dissociate impulsiveness from risk taking, as subjects can only bet large amounts during the increasing trials if they are patient enough to wait for a suitably sized bet to be offered. Output measures include overall proportion bet, delay aversion, deliberation time, risk adjustment, risk taking, and quality of decision making.

One prior study using the CGT in individuals with depression found that, compared to control subjects, depressed subjects accumulated fewer points over the course of the task (mean MDD = 317.1, mean HC = 440.8), took longer to make decisions on a trial by trial basis (mean MDD = 3698.0ms, mean HC = 2484.5ms), and showed reluctance to bet more for safer rounds (i.e. when there were many more of one color block than the other, increasing the likelihood of the token being hidden under that color box) (HC betting proportions for: 6:4=0.41, 7:3=0.57, 8:2=0.74, 9:1=0.82; MDD betting proportions for: 6:4=0.46, 7:3=0.57, 8:2=.65, 9:1=.73) (Murphy, et al., 2001).

Graded Naming Test

The Graded Naming Test (GNT) assesses subjects' semantic verbal memory by asking them to name objects depicted in a series of line drawings. The primary output measure is the percent of correctly identified drawings. This test has frequently been used in studies of individuals with dementia, as reductions in semantic verbal memory (specifically naming of objects) may be an early sign of dementia onset (Blackwell, et al., 2004). The test does not have published results when used in populations with MDD, but was included in our battery for its high test-retest reliability and utility in discriminating changes in subjects' semantic verbal memory over time with great sensitivity (Bird & Cipolotti, 2007), which is important for our ongoing assessment of the short and long term neurocognitive safety of SCCwm DBS for TRD. However, because the long term storage and retrieval of semantic verbal information draws on resources of the hippocampus (McKenna & Warrington, 1980), a region that has been shown to have atrophy in individuals with depression (which is greater with longer duration of illness)

(Sheline, et al., 1996), it is reasonable to expect that we may see impaired performance in this test in the depressed groups relative to the healthy control group, with a greater deficit in the TRD group.

Intra/Extra-Dimensional Set Shift

The Intra/Extra-Dimensional Set Shift (IED) is primarily a test of the cognitive domain of attentional set shifting, which requires the ability to "shift" one's attention from one set of task rules to another. This test is similar to the Wisconsin Card Sorting Test (WCST) (Heaton, Chelune, Talley, Kay, & Cortiss, 1993). Subjects are presented with two figures and asked to determine which of the two is "correct." Subjects must learn, by trial and error, which shape is currently the "correct" one, and after they have proven that they have learned this by selecting the "correct" shape six times in a row, the computer changes the stimuli presented and a new "correct" shape must be identified. The early trials are "intra-dimensional" because the figures are all solid pink "simple" shapes. Later, white lines are superimposed on top of the pink shapes, and the figures presented now become "compound" shapes because the subject must determine whether the pattern of the superimposed white lines or the underlying pink shape determines the correct figure. These stages thus require the subject to make "extra-dimensional" (ED) (i.e. from pink shape to white lines) shifts in their selective attention. Output measures include stages completed, pre-ED shift errors, ED shift errors, total errors, and adjusted errors (corrected for number of stages completed).

Individuals with depression have frequently shown deficits in the IED compared to controls (Purcell, et al., 1997; Purcell, Maruff, Kyrios, & Pantelis, 1998; Taylor

Tavares, et al., 2007). In these studies, individuals with depression, for example, completed fewer stages of the IED (i.e. they were capable of learning fewer rules) than did healthy controls (mean stages completed: MDD = 7.9, HC = 8.7 ($p=0.02$); percentage of subjects to pass all nine stages: MDD = 50%, HC = 85% ($p<0.01$) (Purcell, et al., 1997)).

Stockings of Cambridge

The Stockings of Cambridge (SOC) tests for spatial planning, with emphasis on executive functioning and working memory. The subject is presented with two rows of colored balls hanging in "stockings" and instructed to rearrange the bottom row of balls so that it matches the top row exactly. Subjects are instructed to plan out their sequence of moves before beginning moving the balls and to try to get the problems right on the first attempt. Subjects are also told the number of moves the entire problem should require (ranging from 1-5). The scored problem set begins with easy (two-move) problems and increases in difficulty, concluding with the most difficult (five-move) problems. There is also a motor control portion of the test to assure any generalized motor slowness in the subject is not accounting for observed differences in time spent planning moves. Output measures include total number of problems solved in the minimum number of moves, as well as, for each level of problem difficulty (2-, 3-, 4-, and 5-move problems), mean number of moves made per problem, mean initial thinking time (time before first ball is moved), and mean subsequent thinking time (sum of all pauses at any point after the first ball is moved).

Performance on the SOC, like other measures of executive functioning, has been found to be poorer in individuals with depression (Michopoulos, et al., 2008). Specific output measures that have been found impaired in depressed individuals compared to healthy controls include the total number of problems solved in the minimum number of moves possible (mean MDD = 6.1, mean HC = 7.6 (p=0.01)), and the amount of subsequent thinking time required in completing the most difficult (5-move) problems (mean log 5-move subsequent thinking time: MDD=7.8, HC=7.2, (p=0.04)).

Verbal Recognition Memory

The Verbal Recognition Memory task (VRM) tests for verbal working memory ability. The subject is instructed to remember a series of words appearing one at a time on a computer screen. The instructions emphasize that the subject should attempt to remember as many of the words as possible, but that remembering the order in which the words appear is not important. The subject is also asked to read each word aloud as it appears on the screen so that the experimenter knows that the subject is attempting to remember the correct word. After the entire word list has been presented, the subject is tested for their memory of the word list in two ways. First, the subject is asked to recall from memory as many words as he or she can remember. Recorded outputs are number of correct words listed and number of novel words (words not part of the original set viewed). Next, the subject is presented with another series of words, half of which are words from the initial word list (targets) and half of which are not (distracters), and asked to say for each whether or not that word was part of the original list the subject was asked

to remember. For this recognition trial, the recorded outputs are number of total correct responses, correct targets, and false positives.

Similar to the GNT, the VRM has not been used in studies of individuals with depression. However, because other tests of verbal memory have suggested impairments in free recall in depressed patients (Selective Reminding Test – Sum of Free Recall: MDD = 63.0, HC=71.4 ($p<0.001$)) (Fossati, Coyette, Ergis, & Allilaire, 2002), it was reasonable to expect that our use of the VRM would provide similar results.

Neuropsychological Data Analysis

Data analysis was performed using the Statistical Package for Social Sciences Version 18 (SPSS, Chicago, IL) running on an Apple MacBook Pro (Cupertino, CA) running Mac OSX version 10.6.3. Demographic data were compared between the HC and MDD groups using an independent samples t-test with alpha of .05 considered significant. Measures with significant differences between MDD and HC groups were then co-varied in the comparisons of neuropsychological data between these groups. However, no demographic characteristic was significantly different between these groups, so independent samples t-tests were used to compare neuropsychological data between groups. For these analyses as well, an alpha of .05 was considered significant, however measures under .10 but above .05 were reported as trends. No correction was made for multiple comparisons.

Next, demographic and clinical data were compared between the TND and TRD groups, and between the TRD UP and TRD BP groups using independent samples t-tests, with alpha of .05 considered significant. Any measure found significantly different

between these groups was then entered as covariates in the comparison of neuropsychological data between groups. Because significant clinical differences were found in both of these comparisons, analysis of covariance (ANCOVA) was used to compare each neuropsychological output measure between groups, with that measure as the dependent variable, patient group as the fixed factor, and clinical/demographic variables identified as significant entered as covariates. Again, an alpha of .05 was considered significant, and an alpha under .10 but above .05 was considered a trend. Again, no correction was made for multiple comparisons.

Table 1. Description of Neurocognitive Tests and Expected Results

Neurocognitive Test	**Brief Description of Task & Implicated Brain Regions**	**Expected Result**
North American Adult Reading Test (NAART)	Subject reads a list of words aloud and pronunciation of each word is scored for accuracy (Control for pre-morbid IQ)	No differences expected between depressed and control groups, as IQ deficits have not previously been observed in depression
Motor Screening Test (MOT)	Subject touches a series of X's appearing on screen in order to orient himself with the touch screen computer (Cerebellum, motor cortex)	This test is used for subject training and gross motor deficit screening only Data will not be analyzed between groups
Affective Go/No-Go (AGN)	Subject sees a series of positive and negative words quickly appearing on screen and must press button for one type only (positive or negative), which alternates (Anterior, Subgenual, and Dorsal Anterior Cingulate; Orbitofrontal cortex)	Depressed groups should be faster to identify negative words than positive words, while the opposite should apply to the control group, reflecting established affective biases in these groups
Cambridge Gambling Task (CGT)	Subject chooses whether token is hidden under a red	Depressed groups should show slower

	square or a blue square based on odds presented and then wagers number of points based on confidence of decision (Orbitofrontal Cortex)	deliberation time and accumulate fewer points compared to control group
Graded Naming Test (GNT)	Subject must identify a series of line drawings depicting common objects (e.g. kangaroo) (Hippocampus)	Depressed groups should identify fewer images than controls due to impairment in semantic verbal memory
Intra/Extra-Dimensional Shift (IED)	Subject learns a series of rules in order to correctly select between two visual patterns (DLPFC)	Depressed groups should complete fewer stages than control group due to deficits in set-shifting
Stockings of Cambridge (SOC)	Subject must arrange balls to match target image, testing spatial planning and executive functioning (DLPFC)	Depressed groups should solve fewer puzzles and require greater planning time for each due to executive functioning deficit
Verbal Recognition Memory (VRM)	Subject must briefly remember a list of non-emotional words presented one by one on screen (tested first for free recall and then for recognition)	Depressed groups should have reductions in free recall due to working memory deficit

	(DLPFC)	
Stroop Task	Subject first reads list of words aloud as fast as possible to gauge verbal processing speed, then reads conflicting color word-pairs to gauge executive functioning (Dorsal Anterior Cingulate Cortex, DLPFC)	Depressed groups should have overall reduction in speed of list reading (reflecting slowed verbal processing speed) and increased interference effect (reflecting executive functioning deficit)

Imaging Protocol

Imaging Parameters

All participants underwent both an anatomic T1 image and a resting Blood Oxygenation Level Dependent (BOLD) fMRI. All subjects were scanned in the Biomedical Imaging Technology Center (BITC) of Emory University Hospital on a 3.0T Siemens Magnetom TIM Trio scanner (Siemens Medical Solutions USA; Malvern PA, USA) with maximum gradient strength of 40 mTm^{-1} using a 12-channel head matrix coil. High-resolution anatomic images were acquired at 1x1x1 mm^3 resolution with an MPRAGE sequence (FOV 224x256x176 mm^3, TR 2600 ms, TE 3.02 ms, FA 8°, GRAPPA factor 2).

Resting state fMRI data was acquired with the Z-SAGA (Heberlein & Hu, 2004) sequence to minimize susceptibility artifacts. One hundred and fifty functional volumes were acquired in thirty 4-mm axial slices (TR 2920 msec, TE1/TE2 30 msec/66 msec, FA 90°, 64 x 64 matrix, in-plane resolution 3.44x3.44 mm2). Physiologic measures (heart rate and pulse) were measured throughout the resting scan. These data were used in the data analysis to correct for physiological noise using RETROICOR (Glover, Li, & Ress, 2000). For resting state functional acquisition, subjects were instructed to clear the mind of any specific thoughts, attempt to think of nothing specifically, and passively view a fixation cross (to discourage eye movement and help prevent subjects from falling asleep). Compliance was assessed during an exit interview.

Image Preprocessing

Imaging data analysis began with preprocessing of the anatomical image (T1) and its normalization to MNI space using AFNI (Cox, 1996). Each subject's registered anatomical image to MNI space was individually assessed for quality of spatial normalization. The average of the T1s from all subjects was also used to determine the quality of the overall registration. Each subject's image was then automatically segmented into three parts (white matter, grey matter, and cerebrospinal fluid (CSF)) using a tool from the Analysis Group at the Oxford Centre for Functional MRI of the Brain (FMRIB), FMRIB's Automated Segmentation Tool (FAST) (Zhang, Brady, & Smith, 2001). This was performed to further remove noise artifacts from the resting fMRI dataset further explained below.

Next, the resting BOLD fMRI was slice time corrected and also then motion corrected. Functional data was then normalized into standard MNI space for all subsequent analyses using registration parameters determined by the normalization of the T1 image to MNI space.

FC Seed ROI Selection

Seed ROI's were selected based on the specific group differences in neuropsychological performance identified from comparisons of 1) all patients (both treatment naïve and treatment resistant) versus healthy controls and 2) treatment naïve versus treatment resistant depressed individuals. The results of these comparisons are discussed at length in the results section, but select significant measures are described here briefly for the purpose of justifying seed ROI selection.

The first ROI was selected based on a difference in affective bias for negative versus positive words that was identified between the combined depressed (TND plus TRD) and control groups on the AGN, with control individuals slightly faster to identify positive target words than negative target words, and depressed individuals slightly faster to identify negative target words than positive target words. Elliott et al. (2002) scanned healthy and depressed individuals with fMRI while they performed a slight variant on this task (the version employed in the current study used only emotional –i.e. positive and negative– words for targets and distracters, whereas their version also included conditions with neutral word targets and distracters). They identified a region in the right anterior cingulate cortex (Talairach coordinates 6, 36, 15) that showed greater activation in depressed individuals when they were responding to words of negative valence and greater activation in the control group when they were responding to words of positive valence. Based on this study, we defined an ROI (referred to as ELLIOT in this thesis) with 6mm radius and center of mass located in their reported coordinates (converted to MNI space) of peak activation difference.

The second ROI was selected based on the finding of psychomotor retardation (PR) evident in the combined depressed group, and worse in the TRD group than the TND group. This is based on significantly slowed Stroop word, color, and color-word combination scores in the combined depressed group versus controls, and highly significantly slowed AGN positive and negative target word identification latencies in the TRD group compared to the TND group. Videbech (2002) correlated resting PET with a number of depressive symptom domains and determined that PR was most associated with the dorsolateral prefrontal cortex (Videbech, et al., 2002). Additional support for a

DLPFC ROI comes from the slight deficits observed in a number of the tasks examining executive functioning, a cognitive domain also believed to be mediated primarily by the DLPFC and that commonly is associated with abnormal DLPFC activity in imaging studies of executive task performance in individuals with depression (Thomas & Elliott, 2009).

Imaging Data Analysis

The averaged time courses for the defined ROI seeds were extracted from the preprocessed functional imaging data. Each ROI time course was then correlated with the time course of every other voxel in the brain to determine which areas of the brain were "functionally connected" with the seed regions. The resulting FC correlation maps were then converted using Fisher t-to-z transform such that parametric tests between groups could be performed. FC maps were compared across groups using independent samples t-tests. Comparisons included 1) All depressed individuals vs. healthy controls, 2) TND vs. TRD patients, and 3) TRD UP vs. TRD BP patients. In addition, ROI FC maps were also correlated with behavioral results to assess relationships between localized BOLD response and behavioral data.

RESULTS

Demographic and Clinical Data

Subject Demographic Information

Table 2. shows a summary of basic demographic information for each group. Independent samples t-tests were performed for each demographic characteristic, between each of the subject group pairs to be tested for differences in neuropsychological data in the subsequent sections (HC vs. MDD, TND vs. TRD, TRD UP vs. TRD UP). No significant difference in any demographic characteristic was present in any of the between group comparisons performed. BDI-II scores confirmed a lack of depressive symptoms in the HC group on the day of neurocognitive testing, and also confirmed that the patient group was significantly greater than the HC group on this measure (mean (SD) BDI-II: HC = 0.8(1.4) MDD = 33.3(11.3), $t=2.79$, $p<.001$).

Patient Clinical Data

Clinical data for the patient groups are summarized in Table 3. Independent samples t-tests were performed to compare data for each clinical variable between the TND and TRD groups, and between the TRD UP and TRD BP groups. Any significant result that emerged from these between-group comparisons was then included as a covariate for all subsequent comparisons (i.e. for each neuropsychological output measure) between those groups. As compared to the TND patients, TRD patients on average had higher scores on both depression rating scales (mean (SD) Baseline 17-Item HDRS: TND= 18.3(3.0) TRD= 23.8(3.2), $t=4.72$, $p<.001$; mean (SD) BDI-II: TND=27.0(8.2) TRD=37.7(11.3), $t=-2.79$, $p=.010$), earlier age of illness onset (mean (SD) TND=30.3(12.7)

Table 2. Group Demographic Data

Characteristic	HC	MDD	T-Test	TND	TRD	T-Test	TRD UP	TRD BP	T-Test
Group N (M, F)	15 (9, 6)	29 (14, 15)	χ^2 = N.S.	12 (7,5)	17 (7,10)	χ^2 =N.S.	10 (3, 7)	7 (4, 3)	χ^2 =N.S.
Age	36.0(10.7)	40.0(10.5)	t = -1.2, p= .237	37.4(12.3)	41.9(9.0)	t=-1.13, p=.268	39.9(9.5)	44.7(8.1)	t=-1.09, p=.293
Years of Education	17.3(2.2)	15.7(2.7)	t= 1.97, p=.055	14.75(2.0)	16.4(2.9)	t=-1.64, p=.111	16.1(2.7)	16.7(3.5)	t=-.413, p=.685
NAART FSIQ	112.1(6.1)	111.5(7.8)	t= .248, p=.806	108.5(9.7)	113.4(5.9)	t=-1.68, p=.106	113.8(7.0)	113.0(4.6)	t=.259, p=.800
Test-Day BDI-II	**0.8(1.4)**	**33.3(11.3)**	**t = -15, p<.001**	**27.0(8.2)**	**37.7(11.3)**	**t=-2.79, p=.010**	35.8(12.3)	40.4(10.1)	t=-.820, p=.425

NAART data missing for 3 HC subjects due to recording device malfunction

BDI-II=Beck Depression Inventory-II, HC=Healthy Control, IQ=Intelligence Quotient, NAART FSIQ=North American Adult Reading Test Full Scale Intelligence Quotient, TND=Treatment Naïve Depressed, TRD=Treatment Resistant Depressed

Table 3. Patient Clinical Data

Characteristic	TND	TRD	T-Test	TRD UP	TRD BP	T-Test
Group N (M, F)	12 (7, 5)	17 (7, 10)	χ^2 =N.S.	10 (3, 7)	7 (4, 3)	χ^2 =N.S.
Age	37.4(12.3)	41.9(9.0)	t=-1.13, p=.268	39.9(9.5)	44.7(8.1)	t=-1.09, p=.293
Baseline 17-Item HDRS	**18.3(3.0)**	**23.8(3.2)**	**t=-4.72, p<.001**	23.4(3.4)	24.4(3.2)	t=-.631, p=.538
Test-Day BDI-II	**27.0(8.2)**	**37.7(11.3)**	**t=-2.79, p=.010**	35.8(12.3)	40.4(10.1)	t=-.820, p=.425
Age of Illness Onset	**30.3(12.7)**	**18.6(8.2)**	**t=3.03, p=.005**	18.1(6.6)	19.4(10.7)	t=-.332, p=.745
Total Number of Depressive Episodes	**1.3(.89)**	**7.0(9.2)**	**t=-2.12, p=.022**	3.2(2.5)	12.4(12.6)	t=-1.92, p=.101
Duration of Present Episode (wks)	**381.4 (762.3)**	**280.0 (238.6)**	**t=.517, p=.609**[§]	**400.0(247.6)**	**108.0(34.6)**	**t=3.68, p=.005**
Duration of Illness (yrs)	**7.1(14.5)**	**23.3(10.8)**	**t=-3.46, p=.002**	21.9(11.5)	25.3(10.1)	t=-.636, p=.534
Lifetime Suicide Attempts	**0.0 (0.0)**	**1.7(3.0)**	**t=-2.33, p=.033**	1.4(1.7)	2.1(4.4)	t=-.488, p=.633
Total Mood Related Hospitalizations	**0.0 (0.0)**	**5.2(5.5)**	**t=-3.90, p=.001**	3.6(2.6)	7.6(7.6)	t=-1.36, p=.217

BDI-II=Beck Depression Inventory-II, HC=Healthy Control, HDRS=Hamilton Depression Rating Scale, TND=Treatment Naïve Depressed, TRD=Treatment Resistant Depressed

[§]Driven by two outliers in the TND Group (With their removal: Mean (SD) TND=93.7(68.0) TRD= 280.0(238.6), t=-3.02, p=.007)

TRD=18.6(8.2), t=3.03, p=.005), a greater number of total lifetime depressive episodes (mean (SD) TND=1.3(.89) TRD=7.0(9.2), t=-2.12, p=.022), a longer lifetime duration of illness (mean (SD) TND=7.1(4.5) TRD=23.3(10.8) t=-3.46, p=.002), and a higher number of past suicide attempts (mean (SD) TND=0.0(0.0) TRD=1.7(3.0), t=-2.33, p=.033) and mood-related hospitalizations (mean (SD) TND=0.0(0.0) TRD=5.2(5.5), t=-3.90, p=.001). Duration of present episode was initially not significantly different between groups (mean (SD) TND=381.4(762.3) TRD=280.0(238.6), t=.517, p=.609), but this was determined to be due to the presence of two outliers in the TND group, and became significant after their removal (new mean (SD) TND=93.7(68.0) TRD= 280.0(238.6), t=-3.02, p=.007). Within the TRD group, only duration of present episode was significantly different between the UP and BP patients (mean (SD) TRD UP =400.0(247.6) TRD BP=108.0(34.6), t=3.68, p=.005).

Neuropsychological Data

A summary of all results can be found in Tables 4 and 5 below. Statistically significant (p=.05) results are marked in bold, and trend (p=.10) results are marked in italics.

Motor Screening Test

All subjects were capable of following commands and completed the MOT. Thus all of the tests below were completed for all subjects unless otherwise noted.

Traditional Stroop Task

Data for the traditional Stroop Task are summarized in Table 4a. Color and Color-Word data for one patient (TRD, BP) were excluded from analysis due to subject color-blindness, but Words data for this subject was included because the words condition does not involve color discrimination.

An independent samples t-test was used to compare HC and MDD performance on each Stroop measure. Significant differences were found for the number of items completed on the Stroop Words (mean (SD) HC=105.3(9.4) MDD=95.6(13.8), p=.019) Stroop Colors (HC=78.5(11.1) MDD=71.1(11.1), p=.045) and incongruent color-word pairs (Stroop Color-Words) (HC=46.9(9.3) MDD=39.8(8.8), p=.017) conditions. These collective differences in each of the speed-based tests suggested that speed needed to be controlled for when comparing Stroop interference effect across groups. Thus raw interference effect (Stroop Colors – Stroop Color-Words) was converted to a percentage of interference on Stroop Color-Words as compared to Stroop Colors by dividing the raw interference effect by Stroop Color score. However, this percent interference was not significantly different between the HC and MDD groups (mean (SD) HC=40.5%(7.6) MDD=44.0%(9.6), p=.233).

Two sets of ANCOVA's were then performed to compare Stroop performance across patient groups. First, TRD patients were compared to TND patients with a separate

Table 4. Regular and Emotional Stroop Task Group Data

4a. Stroop Task

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Stroop Words	**105.3 (9.4)**	**95.6 (13.8)**	**p=.019**	98.7(17.2)	93.4 (10.8)	p=.898	93.9(10.7)	92.6(11.6)	p=.356
Stroop Colors	**78.5 (11.1)**	**71.1 (11.1)**	**p= .045**	71.8(12.1)	70.6(10.6)	p=.174	72.7(10.6)	67.0(10.6)	p=.958
Stroop Color-Words	**46.9 (9.3)**	**39.8 (8.8)**	**p=.017**	40.5(8.7)	39.1(9.2)	p=.176	41.2(8.9)	35.8(9.5)	p=.302
Stroop Interference	40.5%(7.6)	44.0%(9.6)	p=.233	43.8%(7.2)	44.2%(11.2)	p=.729	42.7%(11.8)	46.5%(10.7)	p=.148

Data for 1 TRD BP subject excluded for all measures except stroop words due to color-blindness

4b. Emotional Stroop Task

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Neutral Words	68.8 (9.9)	65.5 (11.1)	p=.353	**65.6 (12.0)**	**65.5 (10.8)**	**p=.049**	70.4 (9.8)	57.6 (7.3)	p=.174
Positive Words	67.2 (13.4)	66.0 (12.3)	p=.766	**66.8(12.6)**	**65.2 (12.4)**	**p=.023**	69.1 (13.8)	58.8 (6.5)	p=.185
Negative Words	68.9 (12.8)	64.9 (10.5)	p=.297	65.5 (11.9)	64.4 (9.6)	p=.209	65.9 (9.5)	62.0 (10.4)	p=.301
Positive Interference	3.0%(7.8)	-0.7%(7.3)	p=.139	-1.9%(4.2)	0.4%(9.3)	p=.251	2.2%(10.8)	-2.4%(6.1)	p=.541
Negative Interference	-0.3%(8.2)	0.5%(8.2)	p=.955	0.07%(5.7)	0.9%(10.3)	p=.177	6.2%(7.4)	-7.4%(8.9)	p=.477

Data missing for 3 subjects (2 TRD UP & 1 TRD BP) for all measures, data for 1 TRD BP subject excluded for color-blindness

HC=Healthy Control, MDD=Major Depressive Disorder, TND=Treatment Naïve Depression, TRD=Treatment Resistant Depression, TRD UP= Treatment Resistant Depression (Unipolar), TRD BP=Treatment Resistant Depression (Bipolar)

ANCOVA for each Stroop measure, using the performance measure as the dependent variable, patient group (TND or TRD) as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates. No significant differences between groups were found. In the second set, TRD UP patients were compared to TRD BP patients using Stroop output measure as the dependent variable, subject group as the fixed factor, and duration of present episode as a covariate. Again no significant differences were observed for any Stroop measure.

Emotional Stroop Task

Data for the Emotional Stroop Task is summarized in Table 4b. Data for one patient (TRD, BP) was excluded from analysis due to color-blindness, and data from three patients (two TRD UP, one TRD BP) were not recorded due to technical difficulties. An independent sample t-test between the HC and MDD groups found no significant group differences for any measure. An ANCOVA for each Emotional Stroop output measure, using the measure as the dependent variable, patient group as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates were performed for to compare the TND and TRD patient groups. This revealed significant differences for the neutral word (mean (SD) TND=65.6(12.0) TRD=65.5(10.8), p=.049) and positive word (mean

Table 5. Cambridge Neuropsychological Test Automated Battery Group Data by Subtest

5a. Affective Go/No-Go

Measure	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Happy Word Latency to ID									
Shift	**463.6 (46.6)**	**546.8 (96.6)**	**p=.003**	**488.3 (81.3)**	**588.2 (86.0)**	**p=.031**	596.7 (106.0)	576.0 (50.9)	p=.793
Non-Shift	**468.8 (64.3)**	**549.4 (81.3)**	**p=.002**	*500.0 (59.5)*	*584.3 (77.4)*	*p=.097*	587.5 (93.2)	579.8 (54.1)	p=.982
Total	**466.1 (52.8)**	**547.9 (87.8)**	**p=.002**	**493.8 (69.8)**	**586.0 (80.1)**	**p=.049**	592.0 (97.8)	577.6 (51.2)	p=.885
Sad Word Latency to ID									
Shift	**478.4 (57.8)**	**536.9 (98.5)**	**p=.041**	**477.6 (72.6)**	**578.7 (94.1)**	**p=.049**	573.1 (95.7)	586.7 (98.8)	p=.860
Non-Shift	*497.2 (72.6)*	*547.4 (88.9)*	*p=.067*	**492.6 (76.2)**	**586.1 (77.5)**	**p=.043**	566.5 (76.5)	614.1 (75.5)	p=.548
Total	**487.8 (60.8)**	**542.6 (90.8)**	**p=.042**	**485.9 (70.8)**	**582.6 (82.8)**	**p=.041**	570.1 (83.8)	600.6 (84.4)	p=.702
Percent Negative Word Bias									
Shift	-2.9%(9.9)	2.0%(10.1)	p=.151	2.0%(7.5)	1.9%(12.7)	p=.677	3.4%(13.8)	-1.0%(11.3)	p=.994
Non-Shift	**-5.9%(8.3)**	**0.6%(8.5)**	**p=.020**	1.9%(6.5)	-0.3%(9.8)	p=.453	3.3%(8.9)	-5.4%(9.1)	p=.392
Total	**-4.4%(5.8)**	**1.1%(8.2)**	**p=.025**	1.7%(3.8)	1.1%(8.2)	p=.993	3.5%(10.5)	-3.4%(9.2)	p=.670
Happy Word Commissions									
Shift	1.7 (1.7)	1.7 (1.7)	p=.884	1.5 (1.4)	1.8 (1.9)	p=.957	1.8 (1.6)	1.7 (2.4)	p=.534
Non-Shift	1.3 (1.4)	1.2 (1.6)	p=.850	1.2 (1.6)	1.3 (1.6)	p=.162	*0.7 (.9)*	*2.1 (2.0)*	*p=.053*
Total	3.1 (2.5)	2.9 (2.9)	p=.848	2.7 (2.6)	3.1 (3.1)	p=.381	2.5 (2.2)	3.9 (4.2)	p=.182
Sad Word Commissions									
Shift	1.7 (2.0)	1.7 (1.9)	p=.927	*2.1 (1.6)*	*1.5 (2.2)*	*p=.095*	**0.4 (0.7)**	**3.0 (2.6)**	**p=.026**
Non-Shift	1.2 (1.3)	1.3 (1.4)	p=.799	1.5 (1.2)	1.2 (1.5)	p=.502	1.2 (1.4)	1.1 (1.7)	p=.303
Total	2.9(3.1)	3.0 (2.8)	p=.858	3.6 (2.7)	2.6 (3.0)	p=.114	**1.6 (1.6)**	**4.1 (3.9)**	**p=.046**
Happy Word Omissions									
Shift	0.7 (0.9)	1.2 (2.3)	p=.361	1.7 (2.8)	0.9 (2.0)	p=.460	1.3 (2.4)	0.4 (1.1)	p=.500
Non-Shift	0.5 (0.6)	1.2 (2.3)	p=.253	1.6 (3.3)	0.9 (1.2)	p=.182	1.0 (1.3)	0.7 (1.1)	p=.295
Total	1.1 (1.2)	2.4 (4.2)	p=.260	3.3 (6.0)	1.8 (2.4)	p=.247	2.3 (2.9)	1.1 (1.3)	p=.926
Sad Word Omissions									
Shift	0.5 (0.9)	1.5 (2.6)	p=.184	**2.5 (3.8)**	**0.8 (0.8)**	**p=.019**	0.8 (0.8)	0.7 (0.8)	p=.554
Non-Shift	0.5 (0.8)	1.1 (2.3)	p=.393	1.4 (3.4)	0.8 (1.1)	p=.336	1.0 (1.2)	0.6 (1.0)	p=.846
Total	1.1 (1.4)	2.6 (4.7)	p=.241	*3.9 (7.1)*	*1.6 (1.5)*	*p=.077*	1.8 (1.6)	1.3 (1.4)	p=.651

5b. Cambridge Gambling Task

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Delay Aversion	0.239 (0.094)	0.223 (0.177)	p=.703	0.299 (0.196)	0.178 (0.152)	p=.255	0.169 (0.138)	0.192 (0.182)	p=.435
Deliberation Time (s)	2.264 (0.859)	2.705 (2.089)	p=.442	2.165 (0.593)	3.023 (2.570)	p=.259	2.344 (0.653)	3.993 (3.889)	p=.331
Overall Proportion Bet	*54.2%(13.5)*	*45.6%(15.0)*	*p=.072*	44.8%(12.7)	46.0%(16.5)	p=.541	45.5%(15.8)	46.8%(18.7)	p=.866
Quality of Decision Making	0.951 (0.090)	0.935 (0.118)	p=.647	0.938 (0.096)	0.933 (0.132)	p=.934	0.950 (0.116)	0.908 (0.159)	p=.981
Risk Adjustment	1.72 (1.21)	1.52 (0.787)	p=.561	1.39 (0.792)	1.60 (0.798)	p=.304	1.85 (.718)	1.23 (0.816)	p=.197
Risk Taking	*0.585 (0.134)*	*0.493 (0.152)*	*p=.058*	0.493 (0.132)	0.494 (0.167)	p=.558	0.493 (0.160)	0.495 (0.189)	p=.982

Data for 2 TND subjects not obtained due to limits in subjects' personal schedules that prevented them from finishing all testing (CGT was picked to skip because it is longest task to complete)

5c. Graded Naming Test

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Percent Correct	65.1%(12.4)	59.9%(11.2)	p=.164	60.3%(13.7)	59.6%(9.5)	p=.916	59.7%(5.3)	59.5(14.1)	p=.628

5d. Intra/Extra-Dimensional Shift

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Stages Completed	8.8 (0.6)	8.4 (1.6)	p=.376	8.8 (0.6)	8.1 (2.0)	p=.541	8.8 (0.6)	7.1 (2.9)	p=.223
Total Errors	18.3 (13.5)	19.3 (13.9)	p= .818	16.8 (9.6)	21.2 (16.3)	p= .772	16.1 (9.6)	28.4 (21.5)	p= .159
Total Errors (Adj.)	20.0 (17.0)	29.7 (40.6)	p=.383	18.8 (14.5)	37.4 (50.9)	p=.595	18.6 (16.3)	64.1 (71.3)	p=.164
EDS Errors	6.7 (9.1)	5.2 (7.5)	p=.581	6.0 (7.5)	4.7 (7.2)	p=.486	5.1 (8.9)	4.1 (6.4)	p=.809
Pre-ED Errors	8.7 (7.3)	10.9 (11.7)	p=.525	7.3 (3.8)	13.4 (14.6)	p=.876	9.3 (6.7)	19.1 (20.9)	p=.305

5e. Stockings of Cambridge

Measure	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Problems Solved in Min Moves	9.3 (2.0)	8.3 (2.0)	p=.134	8.7 (1.8)	8.1 (2.2)	p=.495	8.2 (2.1)	8.0 (2.5)	p=.646
2-Move Problems (Means)									
Moves	2.1 (0.3)	2.1 (0.3)	p=.864	2.1 (0.4)	2.0 (0.0)	p=.997	2.0 (0.0)	2.0 (0.0)	-
Initial Thinking Time	3.003 (3.215)	2.894 (2.416)	p=.900	*2.908 (1.572)*	*2.884 (2.918)*	*p=.071*	3.580 (3.364)	1.891 (1.940)	p=.558
Subsequent Thinking Time	0.337 (0.817)	1.051 (2.345)	p=.148	1.953 (3.379)	0.414 (0.856)	p=.755	0.584 (1.044)	0.171 (0.453)	p=.516
3-Move Problems (Means)									
Moves	**3.0 (0.1)**	**3.3 (0.6)**	**p=.047**	3.1 (0.2)	3.4 (0.7)	p=.898	3.2 (0.5)	3.6 (0.9)	p=.517
Initial Thinking Time	7.900 (5.266)	6.756 (5.416)	p=.506	6.632 (3.390)	6.843 (6.589)	p=.109	*4.200 (2.691)*	*10.619 (8.754)*	*p=.051*
Subsequent Thinking Time	1.030 (2.682)	0.980 (2.332)	p=.949	0.137 (0.248)	1.575 (2.926)	p=.832	0.681 (1.667)	2.852 (3.928)	p=.458
4-Move Problems (Means)									
Moves	5.1 (1.0)	5.4 (1.2)	p=.413	5.5 (1.2)	5.3 (1.3)	p=.350	5.4 (1.3)	5.3 (1.3)	p=.367
Initial Thinking Time	10.144 (6.657)	10.346 (7.657)	p=.931	8.884 (3.959)	11.378 (9.439)	p=.988	11.522 (10.640)	11.174 (8.226)	p=.245
Subsequent Thinking Time	3.571 (6.421)	3.422 (3.470)	p=.921	2.197 (2.126)	4.287 (4.005)	p=.243	3.813 (4.034)	4.963 (4.177)	p=.588
5-Move Problems (Means)									
Moves	*6.2 (1.5)*	*7.1 (1.6)*	*p=.076*	6.7 (1.4)	7.3 (1.7)	p=.535	7.5 (2.0)	7.2 (1.5)	p=.154
Initial Thinking Time	11.632 (6.419)	14.749 (11.62)	p=.341	13.285 (9.185)	15.782 (13.24)	p=.858	17.555 (13.433)	13.263 (13.582)	p=.702
Subsequent Thinking Time	2.263 (4.063)	2.138 (2.669)	p=.903	1.441 (1.828)	2.631 (3.090)	p=1.00	2.592 (2.833)	2.686 (3.664)	p=.809

5f. Verbal Recognition Memory

Characteristic	HC	MDD	T-Test	TND	TRD	ANCOVA	TRD UP	TRD BP	ANCOVA
Free Recall									
Total Correct	*8.9 (2.0)*	*7.7 (2.1)*	*p=.088*	7.9 (1.6)	7.6 (2.4)	p=.589	8.5 (2.1)	6.3 (2.4)	p=.109
Novel Words	0.0 (0.0)	0.1 (0.4)	p= .103	0.0 (0.0)	0.2 (0.6)	p=.819	0.3 (0.7)	0.1 (0.4)	p=.353
Recognition									
Total Correct	23.3 (1.5)	22.9 (1.2)	p=.279	23.2 (1.3)	22.6 (1.2)	p=.289	22.9 (1.2)	22.3 (1.3)	p=.663
Correct Targets	11.5 (1.6)	11.2 (1.0)	p=.566	11.3 (1.2)	11.2 (0.9)	p=.225	11.1 (1.0)	11.3 (0.8)	p=.827
False Positives	0.1 (0.4)	0.8 (2.3)	p=.282	0.2 (0.6)	1.2 (2.9)	p=.290	1.4 (3.7)	1.0 (1.4)	p=.600

HC=Healthy Control, MDD=Major Depressive Disorder, TND=Treatment Naïve Depression, TRD=Treatment Resistant Depression, TRD UP= Treatment Resistant Depression (Unipolar), TRD BP=Treatment Resistant Depression (Bipolar)

(SD) TND=66.8(12.6) TRD=65.2(12.4), p=.023) conditions, but not for the negative word condition or any of the interference measures. The same set of comparisons using TRD UP or TRD BP as the fixed factor and duration of present episode as the covariate showed no significant group differences for any Emotional Stroop measure between those groups as well.

Affective Go/No-Go

Data for the AGN is presented in Table 5a. An independent samples t-test was used to compare performance on each output measure between the HC and MDD groups. Significant differences between these groups were found for time (in milliseconds) to respond to happy words for the shift (mean (SD) HC=463.6(46.6) MDD=546.8(96.6), p=.003), non-shift (mean (SD) HC=468.8(64.3) MDD=549.4(81.3), p=.002), and total (mean (SD) HC=466.1(52.8) MDD=547.9(87.8), p=.002) conditions, and time to respond to sad words for the shift (mean (SD) HC=478.4(57.8) MDD=536.9(98.5), p=.041) and total (mean (SD) HC=487.8(60.8) MDD=542.6(90.8), p=.042) conditions, with the non-shift sad words condition only achieving trend significance (mean (SD) HC=497.2(72.6) MDD=547.4(88.9), p=.067). In addition, a significant difference was found for the measure of percent bias for negative words for the non-shift (mean (SD) HC=-5.9%(8.3) MDD=0.6%(8.5), p=.020) and total (mean (SD) HC=-4.4%(5.8) MDD=1.1%(8.2), p=.025) conditions, but not for the shift condition. Numbers of omission errors and commission errors were not significantly different between groups (for either valence or shift condition).

Next, to compare the TND and TRD groups, an ANCOVA was performed for each AGN output measure, using the measure as the dependent measure, subject group as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates was used to compare the performance of the TRD group to that of the TND group. This revealed a general slowing in the TRD group compared to the TND group on most measures, with significant group differences in time to respond (in milliseconds) to happy words on the Shift (mean (SD) TND=488.3(81.3) TRD=588.2(86.0), p=.031) and Total (mean (SD) TND= 493.8(69.8) TRD=586.0(80.1), p=.049) conditions, with a trend difference on the Non-Shift (mean (SD) TND=500.0 TRD=584.3(77.4), p=.097) condition, and to sad words on all three conditions (Shift mean (SD) TND=477.6(72.6)TRD=578.7(94.1), p=.049; Non-Shift mean(SD) TND=492.6(76.2) TRD=586.1(77.5), p=.043; Total mean (SD) TND=485.9(70.8) TRD=582.6(82.8), p=.041). In addition, the TRD group had a significantly smaller number of sad words omission errors for the shift condition (mean (SD) TND=2.5(3.8) TRD=0.8(0.8), p=.019) with a trend for a smaller number for the total condition (mean (SD) TND=3.9(7.1) TRD=1.6(1.5), p=.077), as well as a trend difference in sad words shift condition commissions (mean (SD) TND=2.1(1.6) TRD=1.5(2.2) , p=.095). No difference was found for negative word bias.

Finally, to compare the TRD UP and TRD BP groups, an ANCOVA was performed for each AGN measure, with the measure as the dependent variable, subject group as the fixed factor, and duration of present episode as a covariate. This comparison did not show significant results for any response latency measure (including bias for negative words),

but did show significant differences in sad words commissions for the shift (mean (SD) TND=0.4(0.7) TRD=3.0(2.6), p=.026) and total (mean (SD) TND=1.6(1.6) TRD=4.1(3.9), p=.046) conditions, as well as a near-significant trend difference in happy words commissions for the non-shift condition (mean (SD) TND=0.7(0.9) TRD=2.1(2.0), p=.053).

Cambridge Gambling Task

Results from the CGT are displayed in Table 5b. Data for two subjects (both TND) were not obtained due to time constraints (subject arrived late and did not have time to complete full battery so CGT was omitted because it is the longest of the CANTAB subtests). First an independent samples t-test was used to compare the HC and MDD groups on each CGT output measure. No measure achieved statistical significance, but trend group differences were observed for overall proportion bet (mean (SD) HC=54.2%(13.5) MDD=45.6%(15.0), p=.072) and risk taking (mean (SD) HC=0.585(0.134) MDD=0.493(0.152), p=.058), both of which were lower in the MDD group.

Next, to compare the TND and TRD groups, an ANCOVA was performed for each CGT output measure, with the output entered as the dependent variable, patient group (TRD or TND) as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates. No group differences were found for any measure. Finally, to compare the TRD UP and TRD BP groups, an ANCOVA was performed for each

CGT output measure, with the output measure as the dependent variable, subject group (TRP UP or TRD BP) as the fixed factor, and duration of present episode as a covariate. Again, no group differences were found for any measure.

Graded Naming Test

Results from the GNT are summarized in Table 5c. An independent samples t-test between the HC and MDD groups was performed to compare the percent of items named correctly between groups. This was not significant. An ANCOVA using percent correct as the dependent variable, subject group (TND or TRD) as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates was not significant. Nor was an ANCOVA using percent correct as the independent variable, TRD UP or TRD BP as the fixed factor, and duration of present episode as a covariate.

Intra/Extra-Dimensional Shift

Data for the IED is presented in Table 5d. Independent samples t-tests were used to compare each output measure between the HC and MDD groups. No significant differences were observed for any of the IED measures. Similarly, ANCOVA's between the TND and TRD patients with each IED measure as a dependent variable, subject group as a fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates did not

reveal significant differences between these groups. Finally, the same set of ANCOVA's was performed again using TRD UP or TRD BP subject group as the fixed factor and duration of present episode as a covariate and did not show a significant group difference on any measure.

Stockings of Cambridge

SOC data are summarized in Table 5e. The first set of comparisons used independent samples t-tests to compare each output measure between the HC and MDD groups. A significant difference was found for mean moves on 3-move problems (HC=3.0(0.1) MDD=3.3(0.6), p=.047), and a trend difference was found for mean moves on 5-move problems (HC=6.2(1.5) MDD=7.1(1.6), p=.076), both higher in the MDD group.

Next ANCOVA's were used to compare each output measure between the TND and TRD groups, using output measure as the dependent variable, subject group as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates. A trend group difference for 2-move initial thinking time was observed (TND=2.908(1.572) TRD=2.884(2.918), p=.071).

Finally, a second set of ANCOVA's were used to compare each SOC output measure between the TRD UP and TRD BP groups, with the output measure as the dependent variable, subject group as the fixed factor, and duration of present episode as a

covariate. A nearly significant trend difference was found for 3-move problem initial thinking time (TRD UP=4.200(2.691) TRD BP=10.619(8.754), p=.051).

Verbal Recognition Memory

Results from the VRM are summarized in table 5f. An independent samples t-test of each output measure between the HC and MDD groups showed a trend for a greater number of total correct words produced during the free recall condition in the HC group (mean (SD) HC=8.9(2.0) MDD=7.7(2.1), p=.088). An ANCOVA for each VRM output measure, using the output measure as the dependent variable, subject group (TND or TRD) as the fixed factor, and baseline 17-Item HDRS, test-day BDI-II, age of illness onset, total number of depressive episodes, duration of present episode, lifetime duration of illness, number of past suicide attempts, and mood related hospitalizations as covariates did not reveal any group differences. Similarly, an ANCOVA for each VRM output measure, using the output measure as the dependent variable, subject group (TRD UP or TRD BP) as the fixed factor, and duration of present episode as a covariate also did not reveal any group differences.

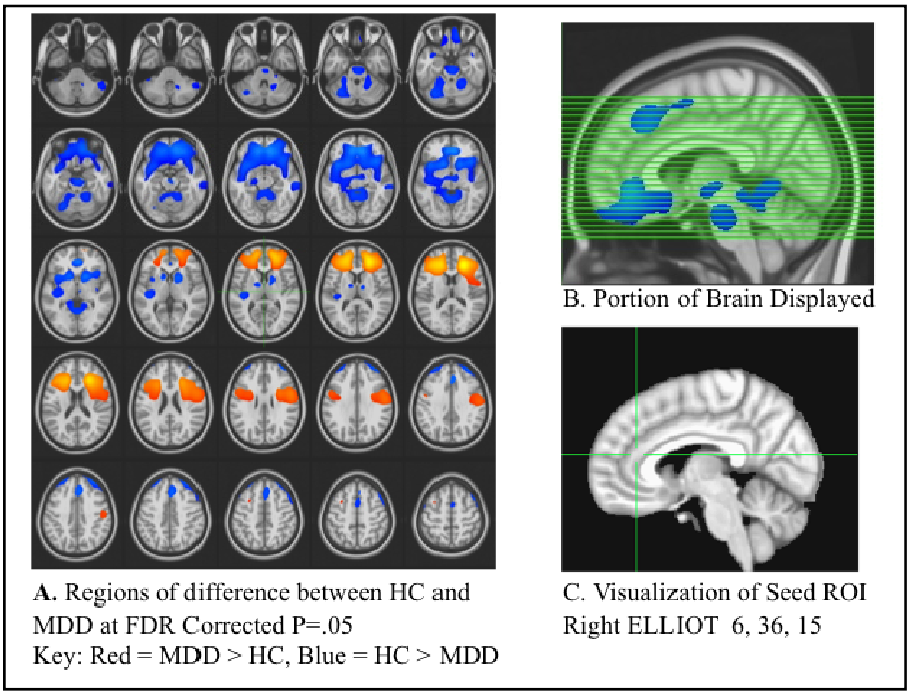
Imaging Results

Due to a very recently discovered major issue with the fMRI data, imaging results could not be obtained in time for inclusion in the thesis (see below for details).

Initial Findings from Group FC Comparison

A whole-brain independent-samples t-test was used to compare the rfMRI FC maps of the HC and MDD groups for the right ELLIOT seed (defined based on previous fMRI study of AGN affective bias (Elliott, et al., 2002)). Results from this comparison are shown in Figure 1. Even using the stringent criteria for significance of p=.05 FDR corrected, large bilateral group differences were found in the insula, orbitofrontal cortex, and caudate, as well as a large block of frontal white matter. These differences were suspicious in their statistical significance, size, and relative symmetry, thus warranting further investigation into possible sources of error.

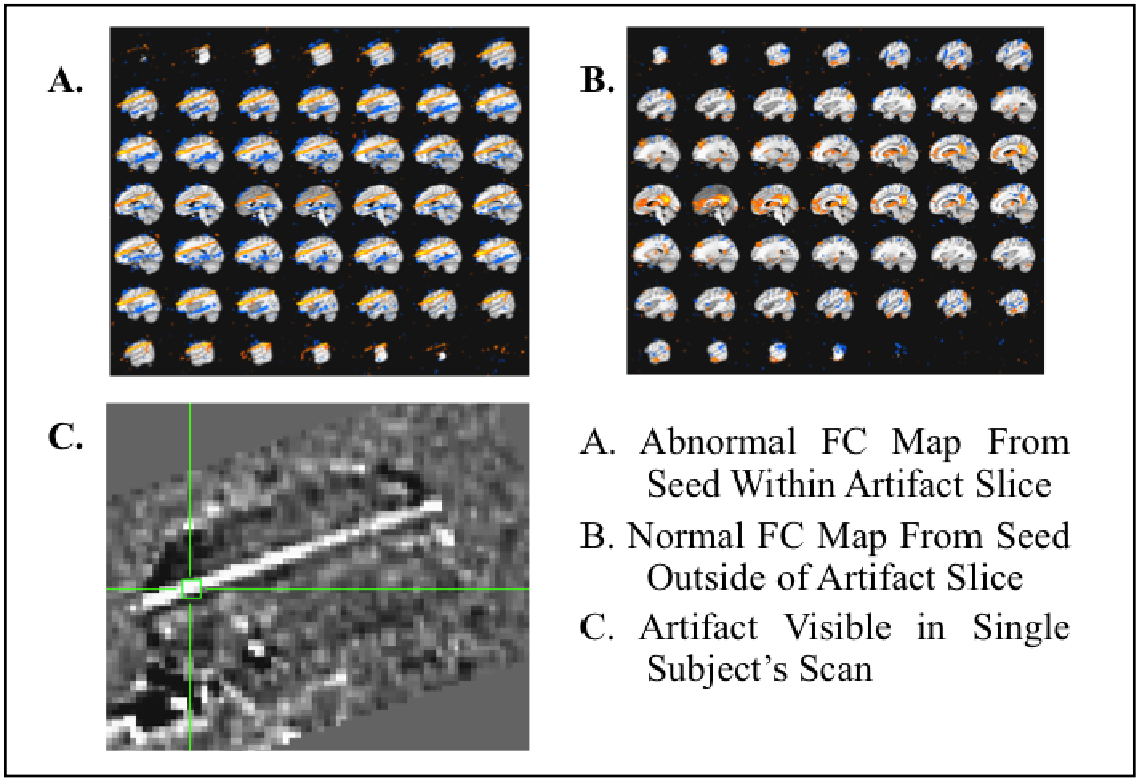
Figure 1. Results from T-Test (HC vs. MDD) of Whole Brain rfMRI FC for Elliot ROI



Subsequent Investigation

Examination of raw data revealed a large artifact in subjects' rfMRI data. This artifact was most prominent in the image slice that included the seed ROI. Previous analyses using these subjects' data had not identified this artifact because these analyses did not use seeds in regions with high correlation to voxels affected by the artifact. Because the current analysis was seeded well within the slice artifact, the regions with which it showed high correlation were simply other voxels within the slice of the artifact. Figure 2. demonstrates these differences. The reasons for this artifact are currently unknown, and are well beyond the scope of this thesis. MRI experts in the Emory BITC are actively working to correct this issue so that analyses using these data may resume.

Figure 2. Comparison of Maps Resulting from Seeds in Affected and Unaffected Slices



DISCUSSION

Summary and Evaluation of Findings

Neuropsychological deficits are an important component of MDD, and a growing body of literature suggests that aspects of illness severity have a large impact on the level of impairment a patient experiences. One aspect of illness severity that has not been explicitly tested previously is that of treatment resistance. While many studies have included patients with recurrent courses of illness and many others have explicitly limited their subject population to medication naïve patients, few studies have sought to dissociate patterns of neuropsychological performance unique to each group. In the present study, we tested the hypothesis that individuals with treatment-resistant depression would show greater deficits on tests of neurocognitive function than would individuals with treatment-naïve depression, while statistically controlling for potentially confounding clinical characteristics.

Differences between All Patients and Controls

First we compared performance on neuropsychological measures between all patients and healthy controls to evaluate the general level of deficits within the study sample of patients. Although we found differences in performance on a number of measures, the most robust differences were in measures of processing speed for both emotional and non-emotional material. Slowed performance was seen on Stroop measures of Words, Colors, and Color-Words. Interestingly, despite these differences, Stroop interference effect was not different between groups, implying that the executive functioning and response inhibition abilities measured by this task were in tact. In addition, although

performance on all measures of the Emotional Stroop (i.e. both those evaluating speed and those evaluating emotional interference) were normal in the patient group, performance on a number of speed related measures on the AGN was reduced, and an emotional bias was present. Interestingly, although the depressed group was slower to respond to both positive and negative words, closer examination of the group statistics reveals that the presence of an affective bias for negative words might be due to a greater amount of slowing (in relation to the control group) for positive word responses relative to negative word responses, as evidenced by a smaller difference in group means, as well as larger p-values, for the negative target words conditions relative to the positive target words conditions (with the difference in Non-Shift negative words condition latencies even failing to achieve significance greater than the trend level). This is consistent with previous literature suggesting that the affective processing bias in depression is due to slowed identification of positive words, rather than faster identification of negative words (Murphy, et al., 1999). However, because a neutral target word condition was not included in the present battery, it is not possible to determine with certainty whether the negative bias in the depressed group was due to a relative slowing of responses to positive words or a relative acceleration of responses to negative words (in relation to the time they would require to respond to neutral words).

Differences between the depressed and control groups were also observed on measures for domains other than processing speed. On the CGT, MDD patients showed trends of lower amounts of overall proportion bet and risk taking, both suggestive of a more conservative betting style. Although previous studies have not identified this type of overall reduction in amount bet (Taylor Tavares, et al., 2007), a failure of depressed

patients to increase their amount bet when presented with an increased odds of winning has been reported (Murphy, et al., 2001). While our HC group's mean overall proportion bet is consistent with previously published data (54.2% vs. 54.0%), our MDD group's average is roughly 12 percentage points lower (45.6% vs. 57.7%) (Taylor Tavares, et al., 2007). Although this study did not report the duration of present episode of their patients, it is possible that the chronic nature of many of our patients' duration of present episode has led to more pronounced anhedonia or greater attitudes of negative future outlook, perhaps making it less rewarding when they experienced a win or causing them to feel less likely to win in general, either way resulting in reduced betting. Although this interpretation is highly speculative and warrants further investigation, it is in line with the conclusions of previous research using this and related tasks (Murphy, et al., 2001).

The final domain showing reduced performance in the MDD group compared to the control group is that of executive functioning and working memory. On the SOC, the depressed group took significantly more moves than the control group on 3-move problems, and showed a trend for requiring more moves than the controls on 5-move problems. One possible explanation for the apparent deficits on these measures but not on the 2-move or 4-move problems is that the 3-move problems are the first level showing a marked increase in difficulty after subjects have just begun to become familiar with the ease of 2-move problems. Thus they are unable to rapidly adjust to this increase, but are then prepared when the shift to 4-move problems occurs shortly thereafter. After two 4-move problems, the subject completes the motor control portion of the test before returning to complete the final two 4-move problems. After these, the subject is presented with 5-move problems, the most difficult. It is thus possible that depressed subjects have

more difficulty adjusting to the rather abrupt increase in difficulty from 4- to 5-move problems, in addition to simply having trouble with this more advanced difficulty level of problems. The other executive task found different between depressed and control groups was a trend decrease in the number of correct words listed in the free recall condition of the VRM. This measure of working memory requires similar use of prefrontal regions to maintain a set of information on-line, though no manipulation is required.

In light of these performance reductions in the MDD group on two tests that capture aspects of executive functioning (SOC & VRM) as well as a task of attention and response inhibition (AGN), it is surprising that impairment was not observed on the IED, which has elements of both attentional set shifting and executive functioning, and has frequently shown poorer performance in this population (Purcell, et al., 1997; Taylor Tavares, et al., 2007). However, a common analysis technique (which was implemented in the studies mentioned above) for the IED is to compare across groups the percentage of subjects that pass the first "Extra-Dimensional" shift condition, the point at which many subjects fail the test. Given the small sample size of the current study, this type of analysis was statistically impractical, as small deviation from the number of subjects expected to pass or fail in either group could easily eliminate a potentially significant result. Thus the most common statistical analysis used for this test could not be implemented, possibly accounting for our failure to identify a deficit on this task.

Effects of Treatment Resistance

The most robust effect of treatment resistance that we identified was that of slowed processing speed in responding to verbal cues, regardless of emotional valence. This was

identified based on ANCOVA's of all neuropsychological measures between the TND and TRD groups (covarying for statistically significant clinical variables between groups). Although no differences were observed on the traditional Stroop task, the TRD group was slightly slower on the Emotional Stroop task for the raw numbers of items completed on the neutral and positive word trials (but not on the positive interference, raw negative, or negative interference measures). On the AGN, the TRD group was significantly slower than the TND group for latency to respond to both happy and sad words on the AGN (though latency for the non-shift happy target words condition showed only a trend difference, presumably because the TRD group did not show increased latency for the non-shift compared to the shift condition, which has been shown in the literature as the norm for control subjects (Murphy, et al., 1999)). The finding of slowed verbal processing speed that was more prominent on a verbal processing task that also required a response to verbal cues (the AGN) than on one that required passive verbal identification (the Stroop tasks) implies an important influence of motoric response speed, and is thus more suggestive of psychomotor retardation than purely cognitive slowing.

The AGN comparisons also showed the seemingly contradictory finding of a simultaneous significant difference in shift condition sad words omissions (with a trend difference for total sad word omissions) and trend difference in shift condition sad word commissions. This implies that TND subjects both fail to respond to sad words during shift sad target words trials more often than do TRD subjects, and also falsely identify sad words as positive words during shift happy target words trials more often than do TRD subjects. This suggests that TND subjects may be slower to learn to start identifying

sad words, as well as slower to learn to stop identifying sad words. The only other trend different measure identified between the TND and TRD groups on the remainder of the tests was that of slightly slower initial thinking time in the TND group on 2-move SOC problems.

In sum, the main finding in the current results is that of an overall slowing in TRD patients' ability to respond to verbal information. Previous literature has established that multiple components of illness severity play a role in determining the type, degree, and severity of the neuropsychological deficits experienced by an individual with depression. These include, but are not limited to, symptom severity (i.e. higher HDRS or BDI-II scores), recurrent vs. single-episode depression, a longer duration of illness, and age of onset (McClintock, Hussain, et al., 2010). However, in the current study, each of these issues was statistically controlled for.

Our results can therefore be interpreted in two ways. Of note, one issue that cannot be dissociated between the TND and TRD groups is exposure of the TRD group to medications, but not the TND group. Thus it is not clear whether an inherent inability of TRD patients to recover from multiple antidepressant treatment trials confers a reduction in psychomotor speed, or if the exposure to multiple treatments and medications comes with the cost of reductions in processing speed. The potential effect of past and present medication exposure is particularly relevant to studies of treatment-resistance, as some of the medications and therapies reserved for treatment resistant cases (e.g. tri-cyclic antidepressant medications, electroconvulsive therapy) can be those with the greatest likelihood of undesirable neurocognitive side effects. In addition, some of the patients in the TRD group were taking benzodiazepines– a medication with known sedative effects

(Buffett-Jerrott & Stewart, 2002)– at the time of testing, which may have affected reaction times on the AGN. While these issues were not examined in depth for this thesis, future analyses with this data will explore the role of medication effects by determining which patients had prior (or current) medication exposures that could have influenced their results.

Effects of Bipolar Disorder

Overall, the TRD BP and TRD UP groups were largely similar for the majority of measures. Differences on the AGN included a higher number of commissions on the sad words shift condition (which led to a higher total commissions number for sad words), and a nearly significant trend increase in commissions on the happy words non-shift condition. The only other difference identified between these patient groups was a nearly significant trend increase in initial thinking time on 3-move SOC problems.

It is noteworthy that the task to show the most differences between the TRD UP and TRD BP groups was the AGN, as that test has shown differences between manic patients and depressed patients (Murphy, et al., 1999). Granted, the bipolar patients in the current study were currently depressed and were of the bipolar II type (thus they had prior hypomanic, not manic, episodes). Still, they showed an increase in commission errors for both valences, which has been seen previously in manic patients (Murphy, et al., 1999). The differential effect of shift condition (group difference was on happy shift condition, but sad non-shift condition) implies an interaction between set shifting and emotional processing that may be worth future investigation. In addition, visual inspection of the TRD BP group's affective bias showed an apparent preference for

positive words in the TRD BP group, rather than negative words as in the TRD UP group; this has also been seen in manic patients (Murphy, et al., 1999). However, this difference did not approach significance and may also confounded by a group difference in duration of present episode.

The effects of bipolar disorder was not the primary aim of the current study, but rather the TRD BP vs. TRD UP comparison was included to explore whether differences between the TRD and TND groups were being driven more by the UP or BP patients in the TRD group. The BP and UP patients did not differ on those measures that were most different between the TND and TRD groups, namely the speed-related components of the AGN. This suggests that the difference in speed between the TRD and TND groups was not due to one of the TRD subgroups more than the other.

Interestingly, a recent large study from France (EPIDEP) suggested that BP II depressive episodes may be associated with less psychomotor retardation than UP MDD, and instead may show a psychomotor activation (Hantouche & Akiskal, 2005). The results from the current study are not consistent with this finding, as no difference was observed between the UP and BP TRD patients. This may be due to the treatment-resistant nature of the depression in the BP II patients in this study, but this interpretation is confounded by the issue of past and present medication exposures, as well as the small sample sizes of the current TRD UP vs. TRD BP comparison. Illness features specific to treatment-resistant depression in patients with bipolar II disorder are an interesting topic for future work, but are beyond the primary aim of this thesis.

Limitations and Recommendations for Future Studies

Limitations

The present study contains a number of limitations that warrant prudent interpretation of our results. The most important limitation is a lack of correction for a large number of multiple comparisons. Given that we analyzed 60 variables in each group comparison, chance alone predicts that for an alpha of .05 for significance and .10 for trend significance, 3 measures in each comparison should be statistically significant, and 3 more should have trend significance. We did not correct for multiple comparisons in the current study because the traditional method of correction (Bonferroni) would have lowered the p-value required for significance to under .01, thus eliminating all measures that were found significant. This would be particularly problematic for later comparisons within the MDD patients (i.e. TND vs. TRD and TRD UP vs. TRD BP), where smaller numbers of subjects as well as inclusion of covariates in analyses of significance already severely limited power to detect differences using traditional uncorrected alpha levels.

Future Directions

A number of recommendations for future studies exploring the role of treatment resistance as a determinant of neuropsychological performance in individuals with depression can be made from the present study. The first is a more robust set of measurements to dissociate different aspects of psychomotor retardation. For example, patients as a whole completed fewer items on the stroop words, colors, and color-word combinations on the Stroop task. These were complemented by significant group differences in AGN response latency for both positive and negative words. However, in

comparing treatment naïve and treatment resistant patients, no differences were observed in any of the regular stroop measures even though significant differences were observed for latency to respond to both positive and negative words on the AGN (and the magnitudes of the latency differences in this comparison were actually greater than were the latency differences in the HC vs. all MDD comparison). This difference implies that depressed patients as a whole may process verbal and visual material more slowly, but only those with treatment resistance may also have motoric slowing. However, this topic needs further clarification. Tests that may be helpful in future studies include a simple reaction time or finger tapping test.

A second recommendation is for the inclusion of an additional group of patients with prior medication exposure. Although statistical correction was made for comparisons within the depressed group to control for the majority of clinical features that can vary across patients (and account for varying levels of neuropsychological impairment), it was not possible to control for the history medication exposure in the TRD group but not in the TND group. This limited the interpretability of our findings, and should be addressed in future studies of neuropsychological deficits associated with treatment-resistance.

A third recommendation is to further investigate differences in neuropsychological functioning in bipolar vs. unipolar depressed patients with treatment resistance. Related to this is the influence of depressive subtypes (e.g. melancholia). A number of patients in the TRD group can be classified as melancholic, and previous research has shown melancholic patients to have more severe deficits (Michopoulos, et al., 2006). However, this effect was not explored in the current study.

Conclusion

To summarize, the present study sought to discern whether treatment resistant depressed patients had a different neuropsychological profile than treatment naïve depressed patients. We used a number of neuropsychological tests that have previously been demonstrated to show impaired performance in individuals with depression. While performance on many of these tests was largely similar between groups, robust differences were found in time required to complete an affective processing task, with treatment resistant patients requiring roughly one tenth of a second more than treatment naïve patients to respond to cues of either positive or negative words. This effect was significant independent of differences in clinical variables between these patient groups. In addition, this slowing appears uniform across unipolar and bipolar II treatment-resistant depression patients.

Limitations of the current study included a small sample size, lack of control for a large number of multiple comparisons, and lack of the necessary control groups to dissociate treatment resistance from medication exposure. In addition to addressing these issues, future studies on this topic should try to address certain domains (e.g. processing speed/psychomotor retardation) more directly and with a more robust set of tests in order to better dissociate the specific impaired components within these complex domains. In addition, greater efforts should be made to quantify the contributions of different clinical variables to variance in specific task performance (e.g. effects of history of suicide attempts on executive task performance), rather than simply controlling for differences in these clinical variables between groups.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders : DSM-IV-TR* (4th ed.). Washington, DC: American Psychiatric Association.

Bird, C. M., & Cipolotti, L. (2007). The utility of the recognition memory test and the graded naming test for monitoring neurological patients. *Br J Clin Psychol, 46*(Pt 2), 223-234.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med, 34*(4), 537-541.

Blackwell, A. D., Sahakian, B. J., Vesey, R., Semple, J. M., Robbins, T. W., & Hodges, J. R. (2004). Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dement Geriatr Cogn Disord, 17*(1-2), 42-48.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test *Clinical Neuropsychologist, 3*(2), 129-136.

Buckner, R. L., Petersen, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *J Neurosci, 15*(1 Pt 1), 12-29.

Buffett-Jerrott, S. E., & Stewart, S. H. (2002). Cognitive and sedative effects of benzodiazepine use. *Curr Pharm Des, 8*(1), 45-58.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res, 29*(3), 162-173.

Craddock, R. C., Holtzheimer, P. E., 3rd, Hu, X. P., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magn Reson Med, 62*(6), 1619-1628.

Crown, W. H., Finkelstein, S., Berndt, E. R., Ling, D., Poret, A. W., Rush, A. J., et al. (2002). The impact of treatment-resistant depression on health care utilization and costs. *J Clin Psychiatry, 63*(11), 963-971.

Dao-Castellana, M. H., Samson, Y., Legault, F., Martinot, J. L., Aubin, H. J., Crouzel, C., et al. (1998). Frontal dysfunction in neurologically normal chronic alcoholic subjects: metabolic and neuropsychological findings. *Psychol Med, 28*(5), 1039-1048.

Elliott, R., Baker, S. C., Rogers, R. D., O'Leary, D. A., Paykel, E. S., Frith, C. D., et al. (1997). Prefrontal dysfunction in depressed patients performing a complex planning task: a study using positron emission tomography. *Psychol Med, 27*(4), 931-942.

Elliott, R., Rubinsztein, J. S., Sahakian, B. J., & Dolan, R. J. (2002). The neural basis of mood-congruent processing biases in depression. *Arch Gen Psychiatry, 59*(7), 597-604.

Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychol Med, 26*(5), 975-989.

Erickson, K., Drevets, W. C., Clark, L., Cannon, D. M., Bain, E. E., Zarate, C. A., Jr., et al. (2005). Mood-congruent bias in affective go/no-go performance of

unmedicated patients with major depressive disorder. *Am J Psychiatry, 162*(11), 2171-2173.

First, M., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders - Patient Edition (SCID-I/P, 11/2002 revision)*. New York: Biometrics Research.

Fitzgerald, P. B., Oxley, T. J., Laird, A. R., Kulkarni, J., Egan, G. F., & Daskalakis, Z. J. (2006). An analysis of functional neuroimaging studies of dorsolateral prefrontal cortical activity in depression. *Psychiatry Res, 148*(1), 33-45.

Fitzgerald, P. B., Srithiran, A., Benitez, J., Daskalakis, Z. Z., Oxley, T. J., Kulkarni, J., et al. (2008). An fMRI Study of Prefrontal Brain Activation During Multiple Tasks in Patients With Major Depressive Disorder. *Human Brain Mapping, 29*, 490-501.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res, 12*(3), 189-198.

Fossati, P., Coyette, F., Ergis, A. M., & Allilaire, J. F. (2002). Influence of age and executive functioning on verbal memory of inpatients with depression. *J Affect Disord, 68*(2-3), 261-271.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci, 8*(9), 700-711.

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med, 44*(1), 162-167.

Goldstein, R. Z., Leskovjan, A. C., Hoff, A. L., Hitzemann, R., Bashan, F., Khalsa, S. S., et al. (2004). Severity of neuropsychological impairment in cocaine and alcohol addiction: association with metabolism in the prefrontal cortex. *Neuropsychologia, 42*(11), 1447-1458.

Gorwood, P., Corruble, E., Falissard, B., & Goodwin, G. M. (2008). Toxic Effects of Depression on Brain Function: Impairment of Delayed Recall and the Cumulative Length of Depressive Disorder in a Large Sample of Depressed Outpatients. *Am J Psychiatry, 165*, 731-739.

Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol, 21*(4), 424-430.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*, 56-62.

Hantouche, E. G., & Akiskal, H. S. (2005). Bipolar II vs. unipolar depression: psychopathologic differentiation by dimensional measures. *J Affect Disord, 84*, 127-132.

Harvey, P. O., Fossati, P., Pochon, J.-B., Levy, R., Le Bastard, G., Lehericy, S., et al. (2005). Cognitive control and brain resources in major depression: An fMRI study using the n-back task. *NeuroImage, 26*, 860-869.

Harvey, P. O., Le Bastard, G., Pochon, J. B., Levy, R., Allilaire, J. F., Dubois, B., et al. (2004). Executive functions and updating of the contents of working memory in unipolar depression. *J Psychiatr Res, 38*(6), 567-576.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. C., & Cortiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Lutz, FL: Psychological Assessment Resources, Inc.

Heberlein, K. A., & Hu, X. (2004). Simultaneous acquisition of gradient-echo and asymmetric spin-echo for single-shot z-shim: Z-SAGA. *Magn Reson Med, 51*(1), 212-216.

Hugdahl, K., Rund, B. R., Lund, A., Asbjornsen, A., Egeland, J., Ersland, L., et al. (2004). Brain activation measured with fMRI during a mental arithmetic task in schizophrenia and major depression. *Am J Psychiatry, 161*(2), 286-293.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 593-602.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 617-627.

Mayberg, H. S. (2006). Defining Neurocircuits in Depression. *Psychiatric Annals, 36*(4), 259-268.

Mayberg, H. S. (2009). Targeted electrode-based modulation of neural circuits for depression. *J Clin Invest, 119*(4), 717-725.

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: A review and synthesis. *Neuropsychology, 24*(1), 9-34.

McClintock, S. M., Hussain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: A review and synthesis. *Neuropsychology, 24*(1), 9-34.

McKenna, P., & Warrington, E. K. (1980). Testing for nominal dysphasia. *J Neurol Neurosurg Psychiatry, 43*(9), 781-788.

Michopoulos, I., Zervas, I. M., Pantelis, C., Tsaltas, E., Papakosta, V. M., Boufidou, F., et al. (2008). Neuropsychological and hypothalamic-pituitary-axis function in female patients with melancholic and non-melancholic depression. *Eur Arch Psychiatry Clin Neurosci, 258*(4), 217-225.

Michopoulos, I., Zervas, I. M., Papakosta, V. M., Tsaltas, E., Papageorgiou, C., Manessi, T., et al. (2006). Set shifting deficits in melancholic vs. non-melancholic depression: preliminary findings. *Eur Psychiatry, 21*(6), 361-363.

Moreines, J. L., McClintock, S. M., & Holtzheimer, P. E. (In Press). Neuropsychologic effects of neuromodulation techniques for treatment-resistant depression: A review. *Brain Stimulation*.

Murphy, F. C., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., Paykel, E. S., et al. (2001). Decision-making cognition in mania and depression. *Psychol Med, 31*(4), 679-693.

Murphy, F. C., Sahakian, B. J., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., et al. (1999). Emotional bias and inhibitory control processes in mania and depression. *Psychol Med, 29*(6), 1307-1321.

Nemeroff, C. B. (2007). Prevalence and Management of Treatment-Resistant Depression. *J Clin Psychiatry, 68*, 17-25.

Paelecke-Habermann, Y., Pohl, J., & Leplow, B. (2005). Attention and executive functions in remitted major depression patients. *J Affect Disord, 89*(1-3), 125-135.

Purcell, R., Maruff, P., Kyrios, M., & Pantelis, C. (1997). Neuropsychological function in young patients with unipolar major depression. *Psychol Med, 27*(6), 1277-1285.

Purcell, R., Maruff, P., Kyrios, M., & Pantelis, C. (1998). Neuropsychological deficits in obsessive-compulsive disorder: a comparison with unipolar depression, panic disorder, and normal controls. *Arch Gen Psychiatry, 55*(5), 415-423.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., et al. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. *J Int Neuropsychol Soc, 4*(5), 474-490.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Stewart, J. W., Nierenberg, A. A., Thase, M. E., et al. (2006). Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med, 354*(12), 1231-1242.

Schlosser, R. G., Wagner, G., Koch, K., Dahnke, R., Reichenbach, J. R., & Sauer, H. (2008). Fronto-cingulate effective connectivity in major depression: A study with fMRI and dynamic causal modeling. *Neuroimage*.

Segal, Z. V., Gemar, M., Truchon, C., Guirguis, M., & Horowitz, L. M. (1995). A priming methodology for studying self-representation in major depressive disorder. *J Abnorm Psychol, 104*(1), 205-213.

Sheline, Y. I., Wang, P. W., Gado, M. H., Csernansky, J. G., & Vannier, M. W. (1996). Hippocampal atrophy in recurrent major depression. *Proc Natl Acad Sci U S A, 93*(9), 3908-3913.

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., et al. (2008). Brain spontaneous functional connectivity and intelligence. *Neuroimage, 41*(3), 1168-1176.

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18*, 643-662.

Taylor Tavares, J. V., Clark, L., Cannon, D. M., Erickson, K., Drevets, W. C., & Sahakian, B. J. (2007). Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. *Biol Psychiatry, 62*(8), 917-924.

Thomas, E. J., & Elliott, R. (2009). Brain imaging correlates of cognitive impairment in depression. *Front Hum Neurosci, 3*, 30.

Trivedi, M. H., Fava, M., Wisniewski, S. R., Thase, M. E., Quitkin, F., Warden, D., et al. (2006). Medication augmentation after the failure of SSRIs for depression. *N Engl J Med, 354*(12), 1243-1252.

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., et al. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry, 163*(1), 28-40.

Vasic, N., Walter, H., Sambataro, F., & Wolf, R. C. (2008). Aberrant functional connectivity of dorsolateral prefrontal and cingulate networks in patients with major depression during working memory processing. *Psychol Med*, 1-11.

Videbech, P., Ravnkilde, B., Pedersen, T. H., Hartvig, H., Egander, A., Clemmensen, K., et al. (2002). The Danish PET/depression project: clinical symptoms and cerebral blood flow. A regions-of-interest analysis. *Acta Psychiatr Scand, 106*(1), 35-44.

World Health Organization. (2008). *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: WHO Press.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging, 20*(1), 45-57.