Distribution Agreement

Micaela McCall April 4, 2018

Analyzing Non-verbal Behavior Throughout Recovery in a Sample of Depressed Patients

Receiving Deep Brain Stimulation

By

Micaela McCall

Andrea Crowell, MD

Adviser

Neuroscience and Behavioral Biology

Andrea Crowell, MD

Adviser

Michael Crutcher, PhD

Committee Member

Robert McCauley, PhD

Committee Member

2018

Analyzing Non-verbal Behavior Throughout Recovery in a Sample of Depressed Patients

Receiving Deep Brain Stimulation

By

Micaela McCall

Andrea Crowell, MD

Adviser

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

Neuroscience and Behavioral Biology

2018

Abstract

Analyzing Non-verbal Behavior Throughout Recovery in a Sample of Depressed Patients Receiving Deep Brain Stimulation

By Micaela McCall

Background. Traditional assessments of depression involve scales based on verbal report. Such scales are used as measures of efficacy in studies of deep brain stimulation (DBS) as a treatment for major depression. These scales lack detailed measurement of non-verbal behavior, despite evidence that non-verbal behavior can provide a window into the unconscious and involuntary physical manifestations of depression. Traditional scales thus may not be sufficient in understanding the recovery of DBS patients as they learn to understand normal negative emotions as transient rather than chronic.

Purpose. This study examines non-verbal behavior in a sample of patients receiving DBS over the first 6 months of treatment. The purpose of this study is to uncover groups of related non-verbal behaviors and investigate the relationship between non-verbal behaviors and verbal self-report scales at different phases of recovery.

Methods. Clinical interviews of twelve DBS patients were analyzed at three time points (1 week pre-operative, ~3 months, and 6 months after start of stimulation), using an ethogram to assess the frequencies of 42 non-verbal behaviors. Beck Depression Inventory (BDI) and Hamilton Depression Rating Scale (HDRS) were also collected at all time points.

Results. Factor analysis groups non-verbal behaviors into three factors: *react, engage/fidget,* and *disengage.* Two-way repeated measures ANOVA shows that scores on the three factors change differently from each other over time. Mixed effects modelling provides evidence that the frequency of non-verbal behaviors related to reactivity and engagement increase as BDI score decreases. Lastly, the non-verbal behavior displayed at ~3 months is more similar to that at 6 months than pre-op.

Conclusion. Non-verbal behavior is a rich source of information about clinical states that cannot be contained by omnibus psychomotor scores on traditional depression rating scales. This study affirms the usefulness of assessment of non-verbal behavior, particularly during transitional phases in which patients re-learn to discriminate transient from chronic negative affect.

Analyzing Non-verbal Behavior Throughout Recovery in a Sample of Depressed Patients

Receiving Deep Brain Stimulation

By

Micaela McCall

Andrea Crowell, MD

Adviser

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

Neuroscience and Behavioral Biology

2018

Acknowledgements

I would like to acknowledge my adviser Dr. Andrea Crowell for her support throughout this project and throughout all my work in this lab. I would like to thank Dr. Helen Mayberg for her encouragement and advice with the development and execution of this project. I would like to acknowledge Dr. Robert McCauley for our stimulating conversations that helped me formulate the concepts around this project, and Dr. Michael Crutcher for his advice and guidance with strategies for data analysis. I would also like to thank Dr. TJ Murphy for his indispensable help in the execution of the statistical analysis in this project.

Table of Contents

Purpose and rationale……………………………………………………………………..................1
Hypothesis…………………………………………………………………………………………… 2
Introduction……………………………………………………………………………………………3
Major depressive disorder and treatment resistance…………………………………………… 3
Neurobiology of depression…………………………………………………………….......... 5
Deep brain stimulation…………………………………………………………………........... 6
Hamilton Depression Rating Score and Beck Depression Inventory…………………………8
Psychomotor symptoms of depression……………………………………………………..... 9
The rough patch………………………………………………………………………………… 11
The analysis of non-verbal behavior…………………………………………………………..12
Ethnography and psychiatry………………………………………………………............... 13

Methods………………………………………………………………………………………… 17
Patients………………………………………………………………………………………….. 17
Surgery and treatment phases…………………………………………………………............... 17
Clinical assessment and time points……………………………………………………….…..... 18
Ethogram…………………………………………………………………………………………18
Collection of non-verbal behavior data………………………………………………………. 19
Data analysis………………………………………………………………………………….... 19

Results………………………………………………………………………………................... 24
Ethogram………………………………………………………………………….................. 24
Factor analysis……………………………………………………………………………….….... 24
Treatment Rating Scores……………………………………………………………………….. 24
Non-verbal behavior factors over time…………………………………………………….….... 25
Association between BDI and non-verbal factor scores………………………………………. 26
Association between HDRS-17 and non-verbal factor scores…………………………….…..... 28
Association between BDI and individual non-verbal behavior scores……………………… 29

Discussion………………………………………………………………………………………… 31
The ethogram…………………………………………………………………………….......... 31
Factor analysis………………………………………………………………………............... 32
Non-verbal behavior factors over time………………………………………………………… 33
Association between BDI and non-verbal factor scores ……………………………............... 34
HDRS-17 and BDI…………………………………………………………………………… 37
The behaviors within the factors……………………………………………………………... 38
Importance for clinical assessment and future directions……………………………………… 40
Limitations…………………………………………………………………………………......... 42

Conclusion ……………………………………………………………………………….......... 44

References………………………………………………………………………………………… 59

Tables and Figures

Table 1. Ethogram………………………………………………………………………… 46

Table 2. Factor scores scaled with mean of zero……………………………………….…. 47

Figure 1. Location of the electrodes implanted during DBS surgery ……………………….. 48

Figure 2. Average BDI scores every 2 weeks for the first 6 months of stimulation………… 49

Figure 3. Correlation plot of non-verbal behavior scores using Pearson Correlations……….50

Figure 4. Diagram of factor loadings generated from confirmatory factor analysis………... 51

Figure 5. Non-verbal factor scores over time……………………………………………... 52

Figure 6. BDI by non-verbal factor score over time……………………………………….53

Figure 7. HDRS-17 by non-verbal factor score over time…………………………………54

Supplementary Table 1. Factor loadings based on exploratory factor analysis………………55

Supplementary Table 2. P-values of pairwise comparisons using paired t-tests…………..... 56

Supplementary Figure 1. BDI by individual behavior scores over time………………….... 57

Supplementary Figure 2. Pause, illustrative gesture, and head to the side scores over time... 58

Purpose and rationale

Dr. Helen Mayberg's lab in the Department of Psychiatry and Behavioral Science of Emory School of Medicine works to develop deep brain stimulation (DBS) of the subcallosal cingulate cortex (Brodmann Area 25) as a treatment for major depressive disorder (MDD). Traditional depression rating scales, such as the Hamilton Depression Rating Scale (HDRS) and Beck Depression Inventory (BDI) are used as efficacy measures of the treatment. These scales are based the patient's verbal report and do not include detailed measures of observable, unconscious, non-verbal behavior. This study involves the quantitative analysis of non-verbal behavior over the first 6 months of DBS stimulation. The first goal of this study is to determine if changes in non-verbal behaviors correlate with changes in traditional depression rating scores between the pre-operative and 6-month time points.

Additionally this study takes an interest in a particular time period, around 3 months after the start of stimulation, which is characterized by a sharp increase in reported symptom severity. It is possible that during this phase, which is referred to as the "rough patch," patients are re-learning to trust the transient nature of normal emotional fluctuations, and they report their distress in terms of the chronic depressive state from which they are emerging. Thus, their high depression scores may reflect confabulation rather than an accurate assessment of their well-being. The second goal of this study is to determine if an analysis of non-verbal behavior can provide new information about the well-being of patients and determine if there is a discrepancy between non-verbal behavior and verbal report at this time point. An analysis of non-verbal behavior of this kind has the potential to reveal the short-comings of rating scales such as the BDI and confirm the importance of observing non-verbal behavior in the assessment of depression.

Hypothesis

I expect that behaviors associated with reactivity and social engagement will increase with decreasing BDI score. During the rough patch, I expect that patients will show wider discrepancy between BDI score and behavior, such that their non-verbal behavior predicts a lower level of depression than what the BDI score indicates.

Introduction

Major depressive disorder and treatment resistance:

Major depressive disorder (MDD) is one of the most prevalent mental health problems in the world. According the World Health Organization (2011), it is the third leading cause of disease burden globally. Using the Disability Adjusted Life Years, which is defined as the "years of life lost due to premature mortality and years of life lost due to time lived in states of less than full health" (WHO), MDD accounts for 4.3% of global disease burden (WHO 2011). Additionally, people with depression are 1.4 times more likely to die of physical health problems than the general population (WHO, 2011).

MDD is diagnosed by having five of nine possible symptoms, which include depressed mood, hopelessness, neurovegetative symptoms, apathy, cognitive deficits, and suicidal ideation (American Psychiatric Association, 2013). MDD is a debilitating disorder that can severely affect quality of life; unfortunately, it has a lifetime prevalence of up to 20% (Kessler, 2003). It is challenging to treat for several reasons, including that patients present with heterogeneous combinations of symptoms and comorbidities with other disorders like Generalized Anxiety Disorder (Ressler and Mayberg, 2007). Thus, there are a large number of clinical phenotypes to address when developing treatments for MDD.

For the latter part of the 20th century, much of depression research centered around the monoamine hypothesis of depression, which proposes that the biological basis of depression lies in a deficiency of the monoamine neurotransmitters serotonin and norepinephrine (reviewed in Hirschfeld, 2000). This line of research has spawned a variety of effective pharmacological treatments that target the balance of monoamines, like serotonin, in the brain (Aberg-Wistedt, 1989; Bunney et al., 1965; Schildkraut, 1965; Carlsson et al., 1957). Despite advances in pharmacological interventions and psychotherapeutic treatments like Cognitive Behavioral Therapy (CBT), many people with major depression remain resistant to these conventional forms of treatment (Neumaier et al., 2016). People with treatment resistant depression (TRD) may undergo more invasive neuromodulation treatments such as repetitive transcranial magnetic stimulation (rTMS) and electroconvulsive therapy (ECT).

Repetitive TMS is an FDA-approved treatment for TRD. It involves passing a pulsed magnetic current through a magnetic coil near the scalp (Rosa and Lisanby, 2012). Different parameters of the current can induce various effects; for instance, different frequencies can be characterized as excitatory or inhibitory (Speer et al., 2000). Meta-analyses of double-blind, randomized, and sham-controlled studies showed that overall, high-frequency rTMS produces statistically significant clinical results when compared to sham treatments (Berlim et al., 2014). However, most people do not remit with this treatment, and some acquire limited benefit in the absence of maintenance treatments (Kedzior et al., 2015; Berlim et al., 2014; Guo et al. 2017). Another option for treating TRD is electroconvulsive therapy (ECT). ECT involves the application of brief electrical pulses to the scalp, causing a seizure (Rosa and Lisanby, 2012). ECT is the most effective treatment for TRD (Ren et al., 2014); however, there are several disadvantages. For example, ECT can cause extensive post-seizure amnesia (Ren et al., 2014; Vasavada et al., 2017). Despite its acute efficacy, nearly 40% of patients experience relapse after ECT (Kellner et al., 2006; Husain et al., 2004). In fact, up to 30% of patients fail to remit after multiple established antidepressant treatments (Rush et al., 2006). Furthermore, the chance of relapse increases as patients experience more depressive episodes. Relapse among patients is >60% after the first depressive episode, >70% after the second, and >90% after the third (American Psychiatric Association, 2013). Due the effect of MDD on quality of life and global disease burden, it is crucial to keep investigating and specifying sub-types of depression and the best ways to treat them.

Neurobiology of depression:

Several lines of research in MDD focus on characterizing the neural circuits behind various symptoms, as well as behind different clinical phenotypes and responses to treatments (Chi et al., 2015; Siegle et al., 2012; Dunlop et al, 2017; Mayberg et al., 2009). Understanding the neurobiology behind depression assists researchers in developing treatments that target these neural abnormalities. However, it has proven difficult to identify the neural circuits underlying depression, which has posed another challenge for the development of treatments for MDD (Ressler and Mayberg, 2007). Likely, the complete neurobiology of MDD is extremely complex. Per Mayberg et al. (2005): “converging clinical, biochemical, neuroimaging, and postmortem evidence suggests that depression is … a systems-level disorder affecting integrated pathways linking select cortical, subcortical, and limbic sites and their related neurotransmitter and molecular mediators” (p. 651). Despite these challenges, several lines of evidence implicate the limbic cortico-striato-thalamo-cortical (CSTC) circuitry in the etiology of MDD (Morishita et al., 2014).

One area of the brain that is particularly implicated in the pathophysiology of MDD is the subgenual anterior cingulate cortex (SCC), otherwise known as Brodmann Area 25 (BA25). BA25 is a key node in distributed CTSC circuits involved in mood regulation and emotional processing; it has direct connections to most of the neural circuits that are abnormal in depression (Holtzheimer and Mayberg, 2011). BA25 is also a principal site of autonomic regulation, playing a role in sleep, appetite, energy (Lane et al., 2013). Functional neuroimaging studies have specifically implicated BA25 activity in acute sadness and depression. In Mayberg et al. (1999), increase in blood flow to BA25 was observed in healthy controls during a sad mood induction. The same pattern was seen in patients with major depression, but the opposite pattern was observed with recovery from depression. Furthermore, Mayberg et al. (2000) showed a decrease in brain glucose metabolism in BA25 associated with SSRI treatment in only those patients who responded to the treatment. The importance of this region is further supported by a meta-analysis of PET and fMRI studies by Sacher et al. (2011), showing increased glucose metabolism in the subgenual and pregenual parts of anterior cingulate cortex in depressed subjects compared to healthy controls. For these reasons, BA25 has been an important target for the development of aggressive treatments for people with TRD that have failed to remit with many or all of the previously mentioned interventions.

Deep brain stimulation:

One such aggressive intervention is deep brain stimulation (DBS) targeted at BA25. Mayberg et al. (2005) described the first open-label study to assess the effectiveness of DBS for TRD. In this study, chronic stimulation of the white matter tracts adjacent to the subgenual gyrus was associated with a sustained remission of depression in four of six TRD patients. The location of the stimulation electrodes is shown in Figure 1 (p. 48). This treatment was made possible partly through the advances in the technology of DBS; it had been used with increasing frequency and success in treating Parkinson's Disease (PD) patients. In PD patients, DBS is used to decrease activity in areas such as the subthalamic nucleus of the basal ganglia, which is characteristically hyperactive in Parkinson's patients (Benabid, 2003; Moro et al., 2010). With chronic DBS stimulation, most Parkinson's patients have experienced significant improvement of motor function (Benabid, 2003; Moro et al., 2010). Similarly, in TRD, DBS decreases activity in the hyperactive BA25 (Lozano et al., 2008). Potential mechanisms of this effect include stimulation of GABAergic afferents or synaptic failure induced by high-frequency stimulation (Ressler and Mayberg, 2007).

Subsequent open-label, uncontrolled studies assessing this intervention have had moderate success. Published studies show around a 50-60% response rate and 30-40% remission rates in the first three years of stimulation (Eitan et al., 2018; Guinjoan et al., 2010; Holtzheimer et al., 2012; Lozano et al., 2008; Lozano et al., 2012; Merkl et al., 2013; Merkl et al., 2017; Puigdemont et al., 2012; Ramasubbu, 2013; Riva-Posse et al., 2017). A study that followed up for 3-6 years showed an average response rate of 64.3 %, as measured at the most recent visit (Kennedy et al., 2011). One multicenter, prospective, randomized trial of BA25 DBS (Holtzheimer et al., 2017) was ended after the initial 6-month blinded, controlled phase because there was no significant difference in the improvement of patients receiving DBS compared to sham stimulation. However, the results after two years were promising, showing a response rate of 53% (Holtzheimer et al., 2017). Together, these results suggest that BA25 DBS for TRD is a promising treatment for patients who remain resistant to many of the available non-invasive treatments.

In Dr. Helen Mayberg's lab in Psychiatry department at Emory School of Medicine, research continues that aims to characterize depression subtypes, understand why patients remit, and develop DBS as a strategy to shift people out of the depressive state (Holtzheimer and Mayberg, 2011). Appropriate patient selection is a key component of these studies. As previously mentioned, MDD can manifest in a variety of symptom combinations. Thus, researchers have attempted to separate depressed patients into distinct subtypes, such as "melancholic" and "atypical," which reflect different combinations of symptoms. However, there is significant overlap between subtypes (Musil et al., 2017), and little evidence that any subtypes predict response to any treatment. Thus, standard inclusion criteria for DBS studies do not mention symptomology (see Methods). Despite this, the clinical team in the Mayberg lab have come to informally look for particular clinical features when selecting patients for DBS, based on trends in the symptoms of past DBS responders. These features include a history of antidepressant response in early episodes, transformation from treatment-responsive to treatment-resistant over time, and lack of emotional reactivity at presentation, which includes slow speech and narrow affective range (Crowell et al., 2015).

Hamilton Depression Rating Score and Beck Depression Inventory:

Official inclusion criteria include a minimum symptom severity score on a standardized depression rating scale, the Hamilton Depression Rating Scale (HDRS-17) (Hamilton, 1960). The HDRS is the primary outcome measure used by the DBS study at Emory. Additional depression scores, such as the Beck Depression Inventory (BDI), are also used as secondary efficacy measures. While the HDRS is an interviewer-rated scale, the BDI is a self-rated scale. Both are highly reliable, valid measures for discriminating depressed from not-depressed patients and tracking recovery in terms of internal and retest reliability (Wang and Gorenstien, 2013; Bagby et al., 2004). However, there are several problematic aspects of these scores. One problematic aspect of the HDRS is that the scoring system of some of the items has not been shown to reflect severity or changes in severity (Bagby et al., 2004). The HDRS has the advantage of being a clinician-rated scale, although it still relies heavily on a patient's self-report of their depressive symptoms.

The BDI, which is a self-rated scale, is more susceptible to patient beliefs, intentions, and placebo effects (Edwards et al., 1984). However, a review by Wang and Gorenstein (2013) determined that it is a sound measure for depression. The BDI measures both somatic and cognitive symptoms, variation in which can be attributed to the severity of depression (Wang and Gorenstein, 2013). The BDI also has good correlation with other depression rating scores, including a correlation of r=0.71-0.75 with the HDRS, according to Wang and Gorenstein (2013). Despite the differences in the rating method of these two scales (interviewer-rated and self-rated), they both share the key commonality of being measured through verbal report of symptoms by patients. Crucially, neither scale includes quantitative measurements of psychomotor or non-verbal symptoms of depression.

Psychomotor symptoms in depression:

By having the primary outcome measure of a treatment be based on verbal report, we miss out on a variety of psychomotor symptoms associated with depression that are not fully captured by these scales. HDRS does include a measure of psychomotor retardation and agitation, meaning slowing and restlessness, respectively; however, single omnibus scores like this lack specificity and can be interpreted in many ways by interviewers. Demonstrating this, Bagby et al. (2004) reports that at least 20% of the time, psychiatrists differed in their ratings of psychomotor agitation and retardation. Similarly, Sobin et al. (1997) states

> motor agitation and retardation can be manifested in multiple motor domains. It is not known which manifestations of psychomotor disturbance result in a positive single-item clinical rating of agitation or retardation, whether most patients simultaneously

> experience more than one motor disturbance, and whether the different manifestations of motor disturbance are equally apparent to observing clinicians (p. 9).

A similar problem is present in the BDI, which, while it does include the somatic symptoms of fatigability and work difficulty, does not directly assess non-verbal/non-conscious psychomotor symptoms of depression. The current study involves an in-depth analysis of various psychomotor behaviors, which aims to broaden the scope of our understanding of the well-being of depressed patients as they go through treatment.

A review of psychomotor symptoms associated with depression by Sobin et al. (1997) emphasizes the importance of analyzing non-verbal behavior in the assessment of depression. Many studies demonstrate distinct trends in non-verbal behavior among depressed patients, which could signal a specific pathophysiology and prognosis (Sobin et al., 1997). A study by Jones and Pansa (1979) reported that depressed patients had higher frequency of self-touching, and lower eye contact, smiling, and eyebrow movement when compared to healthy controls. Ulrich and Harms (1985) performed a factor analysis with videotape-based ratings of non-verbal behaviors and uncovered one retardation and two agitation factors. Reduced speed in decision time and response time in a fixed foreperiod paradigm of a key press task showed motor slowing that is indicative of psychomotor retardation (Ghozlan and Widlocher, 1989). Speech speed is another measure of psychomotor slowing. Depressed patients have increased pause times during counting tasks and clinical interview compared to healthy controls (Sobin et al., 1997). Overall, psychomotor retardation and agitation, as psychomotor symptoms of depression, are not always mutually exclusive. Specific behaviors that fall into either category can be observed among depressed patients and within the same individual (Sobin et al., 1997). This fact demonstrates again that single item retardation and agitation scores in the HDRS are insufficient in assessing the non-verbal behavior associated with depression. This lack in specificity could be obscuring interesting information about the experience and behavior of depressed patients. The current study specifically analyzes non-verbal behavior, a technique that can complement self-report assessments. Such analyses may be particularly helpful during phases of recovery, such as the rough patch, in which the patients' ability to accurately comprehend and communicate their symptoms is questionable.

The rough patch:

In Dr. Mayberg's lab, a group of clinicians and researchers who study DBS manage the care and scientific study of depressed patients receiving this experimental treatment. Throughout all phases of treatment, subjects are closely monitored for changes in their psychological symptoms. Treatment responders typically experience the greatest improvement over the first 6 months of stimulation (Crowell et al., 2015). Typically, treatment responders experience a few weeks of negative affect and disproportionate negative reactivity around 10-12 weeks after the start of stimulation. In the Mayberg lab, this period has been dubbed the "rough patch." During this period depression rating scale scores (like BDI and HDRS) tend to increase as well as fluctuate significantly (See Figure 2, p. 49). According to Crowell et al. (2015), "as patients experience more sustained improvement in mood and more critically, increased emotional range, they move from a state of relative stability around a low negative to relative instability, with heightened emotional sensitivity and reactivity" (p. 4). High scores on traditional rating scales may reflect day-to-day fluctuations in distress rather than the chronic depressive state, which indicates that patients are not necessarily experiencing a relapse of depression. As mentioned in the previous quote, they are rather experiencing an increasingly normal emotional bandwidth, including positive and negative affect. However, they are out of practice discerning normal negative emotions from the depressive state, and so report more depressive symptoms. In responders, this reactivity subsides after several weeks as patients re-learn to trust the transient nature of normal emotional fluctuations (Crowell et al., 2015).

The analysis of non-verbal behavior:

The reason clinicians do not think that the rough patch constitutes a relapse of depression, despite increasing HDRS scores, is simply due to clinical judgement based on interpersonal interactions and clinical interviews. In Crowell et al. (2015), the authors discuss making judgements about the patient's well-being based on "subtle signs of improvement, particularly with regard to patient's reactivity" (p. 3). This points us towards the psychomotor symptoms of depression discussed earlier. As mentioned, DBS patients are typically characterized by psychomotor slowing, slow speech, and limited reactivity; when clinicians notice changes in these psychomotor symptoms, it leads them to believe that something about the patient's depression has shifted, because they are not displaying the same non-reactivity as they previously did.

This situation supports the argument that the HDRS and BDI do not accurately reflect psychomotor symptoms; clinicians who notice such symptoms come to different conclusions about the well-being of the patient than is reflected in the HDRS score. In addition, this situation reveals the short-coming of verbal-report outcomes that they are contingent on how a patient verbalizes/rationalizes their subjective experience (Fiquer et al., 2013). During the rough patch, patients report depression because they do not have experience interpreting *transient* negative emotions. This confounding reflects a classic example of confabulation; people are not sure how to characterize their current experience, so they tell a story that seems to fit based on their previous life experiences. Lastly, these scores are insufficient because they record only conscious processes. Many cognitive processes occur below the level of consciousness and so cannot be verbally reported (Fiquer et al., 2013). Non-verbal behavior is a crucial instance of this; over 60% of human communication is non-verbal, and much of this body language, such as facial expression and posture, is unconscious (Geerts and Brune, 2009).

Ethnography and psychiatry:

Thus, an analysis of non-verbal behavior would add a helpful dimension to any psychological evaluation using verbal report. In this context in particular, an analysis of non-verbal behavior would be helpful in clarifying the state of patients' well-being during the rough patch. One way of analyzing non-verbal behaviors is through ethological methods. Ethology can be conceptualized as the evolutionary study of behavior, and overlaps with psychiatry in its goals of understanding the proximate causes of behavior (Geerts and Brune, 2009). People have argued that psychiatry and ethology are intimately connected. Troisi et al. (1999) argues that psychiatry could benefit from a greater use of ethological methods for several reasons. Firstly, as I have discussed, "several studies have shown that the objective and quantitative recording of patients' behavior may sometimes yield different results from those obtained using rating scales or structured interviews. These findings cast doubt on the validity of routine psychiatric assessments and suggest caution in basing important clinical decisions…exclusively on patients' reports of their symptoms" (Troisi, 1999, p. 906). In addition, ethology builds on evolutionary explanations for behaviors. In people with mental illness, just like in all people, non-verbal communicative behavior reflects an individual's conscious and unconscious strategies for reaching their goals. Analyzing non-verbal behaviors in terms of their significance for the person displaying them *and* their impact on the person receiving them frames an understanding of mental illness as a failure of the patient's strategies to obtain their social goals (Troisi, 1999; Geertz and Brune, 2009). For instance, a patient with MDD may crave sympathy from their partner and dramatize their problems in an attempt to gain this sympathy, which ends up pushing their partner away and exacerbating their depression. Lastly, ethology could lead to the development of a more extensive phenomenology that could be used in the development of more accurate animal models (Troisi, 1999). Therefore, the analysis of non-verbal behavior using ethology can provide different information about patients and open up new modes of analysis in psychiatry.

One might argue that psychiatrists intuitively use ethograms often in assessment and diagnosis. Per Geerts and Brune (2009), "clinicians intuitively use their species-specific endowments for deciphering non-verbal expressions in therapist–patient interactions" (p. 1012). In other words, as humans, psychiatrists are versed in the methods of intraspecific non-verbal communication. One example, the so-called "praecox feeling," experienced by a clinician, is an intuitive recognition of abnormal behaviors related to psychosis in the patient and is regarded as a legitimate tool for assessment (Geerts and Brune, 2009). The case of the rough patch in DBS patients is another excellent example of how intuitive analysis of non-verbal behavior is used by psychiatrists to inform an understanding of the well-being of their patients separately from what information is gained from the traditional depression rating scales.

One important ethological approach is the use of ethograms. An ethogram is a catalogue of discrete species-specific behaviors (Geerts and Brune, 2009). The use of ethograms to systematically analyze non-verbal behavior could bring a dimension of objectivity to the judgements made by clinicians and researchers about the well-being of patients during the rough patch. According to Fossi et al. (1984), “methodologically, it [ethology] could be used as a standardized, objective, and reproducible technique to describe both natural and pathological behavior” (p. 332). In the 1990’s, Bouhuys and Hoodakker (1991) developed an ethogram for use with human subjects during clinical interviews. They used information about the duration and frequency of behaviors, during speaking and listening, to investigate the mutual influences of the patients’ and clinicians’ behavior. Troisi (1999) developed the Ethological Coding System for Interviews (ECSI), which include 37 facial and hand movements. This ECSI contains behavior categories that each “reflect a different aspect of the subject’s emotional and social functioning” (Troisi, 1999, p. 909). It is overly ambitious to claim understanding of what each behavior represents. However, several other researchers have used similar and modified ethograms to analyze depressed patients, both in the hospital and through clinical interviews (Fiquer et al., 2013; Fossi et al., 1984; Bouhuys and Hoofdakker, 1991). Behavior has been analyzed in terms of its prognostic usefulness and its relationship to other evaluations of depression. For instance, Fossi et al. (1984) showed that before treatment, depressed patients in a ward displayed decreased eye-contact, social behavior, and verbal communication, among others, in comparison to after treatment. Bouhuys et al. (1991) found correlations between behaviors associated with speaking effort and restlessness, and clinical measures of activation. Fiquer et al. (2013) used an ethogram to analyze clinical interviews over the course of a neuromodulation treatment. They found that several non-verbal behaviors track with scores on verbal symptom rating scales throughout recovery during treatment.

Similarly to Fiquer et al. (2013) I used a human ethogram to analyze the non-verbal behavior of DBS patients in videos of clinical interviews. I used the information about frequency of various non-verbal behaviors to determine if non-verbal behavior correlates with HDRS and BDI scores over the first 6 months of DBS in this population of patients. Additionally, I used this information to analyze the rough patch and determine if there is a discrepancy in this period between non-verbal behavior and HDRS/BDI scores. This would confirm clinicians' interpretation of this phase and demonstrate the importance of analyzing non-verbal behavior in clinical assessment.

Methods

All studies were approved by Emory University's Institutional Review Board, and I am on the following IRB protocols: IRB00066843, IRB00024883.

Patients:

Twelve patients were drawn from Emory's "Deep Brain Stimulation (DBS) for Treatment Resistant Depression Study." All patients gave written, informed consent for participation. Before participation in the DBS study, patients met DSM IV criteria for major depressive disorder, diagnosed by the Structured Clinical Interview for DSM IV, and met a severity requirement of a score of ≥20 on the HRDS-17. Patients had a current major depressive episode of at least 12 months duration, and failure to respond to at least four different antidepressant treatments. Most previously underwent ECT, some with good response initially that could not be recaptured when symptoms returned. Inclusion criteria were: between the ages of 18-60, no major medical or psychiatric comorbidities such as personality disorders, no substance use disorder in the past 12 months, no active suicidal ideation, no attempt within the past 6 months or more than 2 attempts in the past 2 years, and not pregnant or planning to become pregnant during the study. Patients were allowed to continue medication, which could be changed only if side effects developed.

Surgery and treatment phases:

During bilateral DBS lead placement in the white matter adjacent to BA25, patients were awake, and individual electrode contacts were evaluated based on patient response. Surgery was followed by a 4-week post-operative recovery period during which stimulation was off. Subsequently, patients received 6 months of continuous stimulation during which limited parameter adjustments were made, and no medication changes occurred. Observational follow-up continued for several years and is ongoing, during which changes in medication and psychotherapy are allowed.

Clinical assessment and time points:

Clinical assessment occurred before surgery and weekly through the first 6 months of stimulation. Assessment included administration of the HDRS-17 and BDI as well as other clinical scales, and a clinical interview with a study psychiatrist. Video of the patient's face and upper torso was recorded during each clinical interview. The time-points chosen for this project were: 1 week before surgery (pre-op time point), 12-14 weeks after turn-on (rough patch time point), and 24 weeks after turn-on (6-month time point). Rough patch videos for each patient were selected based on clinical impression rather depression rating scores.

Ethogram:

Behaviors included in this study's ethogram (Table 1, p. 46) were selected from a thorough literature search. Those behaviors that had the most significant and consistent relationship with a traditional outcome score, or were included in several of the reviewed studies, were prioritized for inclusion (Fiquer et al., 2013; Fossi et al., 1984; Troisi, 1999; Bouhuys and Hoofdakker, 1991). The ethogram was based partially on the Ethological Coding System for Interviews, described in Troisi et al. (1999). Inclusion focused on behaviors related to eye, face, head, shoulder, and arm movement, posture, and speech. The ethogram includes precise descriptions of each behavior. When descriptions include time constraints (e.g., pause that lasts at least 3 seconds), numbers were chosen simply to provide a cutoff point and measure of consistency across patients and observers. This ethogram is listed in Table 1 (see p. 46).

Collection of non-verbal behavior data:

The videos were collected from the database and then assigned a random number in Microsoft Excel using its random number generator. In this way, the observer was blinded to the time point at which each video was recorded. The first 8 minutes of the video at each time point was analyzed using the ethogram created for this project. Each 8-minute observation was considered to have started when the clinician asked the first question, eliminating potential shuffling/various movements that occurred before that. The "one-zero" sampling method, described in Troisi et al. (1999), was used to record behaviors described in the ethogram. The 8-minute observations were split into 15-second samples. At the end of each sample, the observer recorded whether each behavior in the ethogram was present in the preceding sample. The recording was often rewound to check the accuracy of the observation. Per Troisi et al. (1999), "one-zero" scores for a behavior are highly correlated to both the frequency of its occurrence and the duration of its occurrences when they are calculated separately.

Data analysis:

Data was analyzed using RStudio. Plots were generated using *ggplot2* package (Wickham, 2009).

Non-verbal behavior scoring:

The score for each behavior is the proportion of the total 32 samples in which the behavior occurred. Therefore, at each time point, each patient received a score for the frequency of occurrence of each behavior during the observation period. Behaviors that occurred in a total of three or fewer videos were discarded and not included in analysis. In addition, groups of similar behaviors that were observed at frequencies slightly above this cutoff were combined into single variables.

Factor analysis:

Factor analysis was used to assess the relationships between behaviors and create groupings of behaviors without *a priori* concepts. Factor analysis reduces a large number of variables to a few theoretical latent dimensions (factors) based on correlations and covariances between variables. Before factor analysis was performed with all non-verbal behaviors, scores were scaled to a mean of 0 and standard deviation (SD) of 1, since some observed behaviors had much wider distributions than others. Additionally, observations for one patient were excluded due to the abrupt onset of an atypical clinical instability and multiple clinical interventions being undertaken just before the 6mo time point.

Exploratory factor analysis (EFA) was performed using only pre-op and 6-month time points, to assess the covariance of behaviors as patients became well. In order to imagine what EFA would look like with a much larger sample, bootstrapping was used to resample 1000 observations from the original dataset, using the *fabricatr* package (Blair, Cooper, Coppock, Humphreys and Rudkin, 2018). A scree plot was constructed to determine the correct number of latent factors for the data, using the *nFactors* package (Raiche, 2010). The scree plot revealed three factors with large eigenvalues. EFA (principal axis factoring method, oblimin rotation) was performed on both the observed sample and the large generated one using the *psych* package

(Revelle, 2017), and the *GPArotation* packages (Bernaards and Jennrich, 2005). Thus, behaviors were reduced down to three latent factors. Factors consisted of non-verbal behaviors with a factor loading of > 0.3 or < -0.3. After the initial EFA, items that did not load significantly onto any factor were eliminated one by one until all remaining items loaded onto one of the factors. The resulting factors were modified slightly using confirmatory factor analysis (CFA, robust maximum likelihood (MLM) estimator), using the *lavaan* package (Rosseel, 2012), on another sample of 1000 observations generated using bootstrapping. Scores for each patient in the original data set on each latent factor were generated through this CFA model. These factor scores were used in subsequent analyses.

Repeated Measures ANOVA:

Two-way analysis of variance for repeated measures (ANOVAr) was performed, using the *ez* package (Lawrence, 2016), to assess the differences between non-verbal scores on the different non-verbal factors and the change over time. Post hoc pairwise t-tests with pooled standard deviation (SD) assessed differences between specific pairs within the ANOVAr predictor groups. Extreme effects were determined using a p-value adjusted for multiple comparisons.

Regression for non-verbal scores and BDI:

Linear regression was used to investigate the association between BDI score and non-verbal factor score as the patients became well (pre-op and 6-month time points). First, multilinear regression was used to model BDI score (dependent) as a function of time and each individual non-verbal factor score, as well as time*score interaction (e.g. BDI=βtime-point + βreact score + βtime-point*react score). Next, linear mixed effects (LME) modelling was used to construct a full model for BDI, using the *lme4* package (Bates, Maechler, Bolker, and Walker, 2015). All the non-verbal factors that were found to be significant in linear regression, as well as their significant interactions with time, were included as fixed effects (e.g. BDI=βtime-point + βreact score + βtime-point*react score + βengage/fidget score + βdisengage score). Also, patient ID was included in the model as a random intercept. ANOVA was used to compare each full LME model with a reduced model to assess the significance of each fixed effect on the fit of the model.

Linear regression was used investigate the association between BDI score and each non-verbal factor score over all three time points. This modelling was conducted to study whether the relationship between non-verbal factor score and BDI score was different at the rough patch time point compared with the other two. The same model was used as mentioned above. LME modelling was used to construct a full model for BDI with all the significant non-verbal factors and their significant interactions with time as fixed effects, as mentioned above. Also, as above, patient ID was included as a random intercept.

Multilinear regression was also used to investigate the association between BDI score and each individual non-verbal behavior scores over time. Then, LME modelling was used to construct a model for BDI including, as fixed effects, all the non-verbal behaviors with a large effect on BDI. As before, a random intercept was included for patient ID. All mentioned regression analyses were performed with non-verbal factor scores and non-verbal behavior scores scaled to a mean of 0.

Regression for non-verbal scores and HDRS-17:

Multilinear regression was used to investigate the association between HDRS-17 score and each non-verbal factor score over time. LME modelling was used to construct a full model for HDRS-17 with all the significant non-verbal factors and their significant interactions with time as fixed effects, and a random intercept of patient ID. These regression analyses were also performed with non-verbal factor scores and non-verbal behavior scores scaled to a mean of 0.

<u>Results</u>

Ethogram:

The behaviors selected from the literature for inclusion in the ethogram for this project are included in Table 1 (p. 46). There were 42 behaviors included. Behaviors that occurred in three or fewer observation points were discarded and not included in analysis. These behaviors were: *frown, intensive body touching, lean forward, light body touching, object touching, object using hand movement, posture shift, relax, settle, and shrug.* The following behaviors were added together into a single variable called *touch head*: *groom, touch hair, touch mouth, touch face*. Again, this was due to low occurrence. In addition, *asymmetric* and *symmetric smile* measurements were added to create a single *smil*e variable due to difficulty reliably discriminating the two. The final ethogram is indicated in bold in Table 1.

Treatment Rating Scores:

Patients who were responders showed a 50% decrease in HDRS-17 score. At the chosen rough patch time points 8/12 patients were responders. At the 6 month time point, 10/12 patients were responders. Remission is defined as an HDRS-17 score of 7 or less. At 6 months, 8/12 patients were characterized as being in remission.

Factor analysis:

Figure 3 (p. 50) shows a correlation plot of all the non-verbal behaviors included in the factor analysis. Factor analysis assessed these correlations to group behaviors into a few latent dimension (factors). After scaling non-verbal factor scores to a mean of 0, Exploratory Factor Analysis (EFA) was performed on non-verbal behavior scores for the pre-op and 6-month time points. A scree plot indicated that three factors should be used. The loadings of each behavior on each factor generated from EFA are reported in Supplementary Table 1 (p. 55). Factor names were chosen based on their constituent behaviors. Factors are: *react, engage/fidget,* and *disengage*. A Confirmatory Factor Analysis (CFA) model was constructed following the results of the EFA with one modification. For the sake of conceptual continuity, *touch head* was moved into the *engage/fidget* factor since it is associated with other fidgeting behaviors. Figure 4 (p. 51) diagrams the factor loadings from CFA on another bootstrapped sample of 1000 observations. Non-verbal behavior outcomes were thus reduced to 3 factors. The CFA model was used to generate factor scores for each patient at each time point, using the original scaled sample. These factor scores are reported in Table 2 (see p. 47).

Non-verbal behavior factors over time:

The mean frequencies of the *react* and *engage/fidget* factors increased over time, and the mean frequency of *disengage* factor decreased over time. The statistical significance of this trend was determined using two-way analysis of variance for repeated measures (ANOVAr), which assessed the differences between non-verbal factor scores over time. This analysis revealed a significant main effect of the non-verbal factor ($F_{2,22}$= 6.66, $p<0.01$) and a significant interaction between time and the non-verbal factor ($F_{4,44}$= 2.78, p=0.038). This means that there are significant differences between the non-verbal factors in the way their scores change over time. Non-verbal factor scores over time are shown in Figure 5 (see p. 52). Post-hoc pairwise t-tests with pooled SD showed that at the rough patch time point, *react* score is significantly higher than *disengage* score (p=8.7e-4, paired t-test) and *engage* scores is significantly higher than *disengage* score (p=0.0021, paired t-test). Within the other two time points, none of the factor scores were significantly different from each other. The post-hoc t-test p-values are shown in Supplementary Table 2 (see p. 56).

Association between BDI and non-verbal factor scores:

All regression analyses were performed with factor scores scaled to a mean of 0. Linear regression was used to model the association between BDI score and score on each non-verbal factor at the pre-op and 6-month time points. For the model of BDI with the predictor variables of *react* score, time, and the interaction between time and *react* score ($F_{3,19}$=9.74, p=<0.01), the model revealed main effects of time (β=-21.91, p<0.01) and *react* score (β=9.66, p=0.049), and a significant interaction between time and *react* score (β=-11.27, p=0.046). The regression model of BDI as a function of time and *engage/fidget* score ($F_{2,20}$=11.51, p<0.01) only produced a significant main effect of time (β=-18.97, p<0.01). The regression model of BDI as a function of time and *disengage* score ($F_{2,20}$=10.45, p<0.01) also only produced a significant main effect of time (β=-19.55, p<0.01). See Figure 6 (p. 53). Thus, only *react* score and time are significantly related to BDI scores between pre-op and 6 months.

Linear mixed effects (LME) modelling was used to investigate the association between BDI and all non-verbal factor scores at pre-op and 6 months. BDI was the outcome variable, and all non-verbal factor scores, time point, and the interactions between time and non-verbal factor scores were included as fixed effects. Patient ID was included as a random intercept. Model comparison using ANOVA showed that the time by *react* score interaction significantly improved the model for BDI (X^2(1)= 4.95, p= 0.026). The slope of the regression line between *react* score and BDI decreases by a coefficient of -11.27±4.80 (standard errors (SE)) between the pre-op and 6-month time points. In other words, between pre-op and 6 months, the magnitude of the negative correlation between the two scores increases. These results confirm that only *react* score and time are significantly related to BDI between pre-op and 6 months.

For the following analyses, rough patch data was included to assess whether it changed the models created with only pre-op and 6 month data. Linear regression was also used to model the association between BDI score and non-verbal factor score over *all time points*. Adding rough patch data changed the model for two time points only minimally; therefore, all three time points were used in all subsequent analyses. The regression model of BDI as a function of time, *react* score, and their interaction ($F_{5,30}$=6.66,p<0.01) produced a significant interaction between time and *react* score in the pre-op to 6-month time interval (β=-11.71, p=0.046) and in the pre-op to rough patch time interval (β=-17.91, p<0.01). Two more regression models for BDI, which were the same as the previous but replaced *react* score with *engage/fidget* or *disengage* scores, produced only significant main effects of time, as was observed in the regression with only pre-op and 6-month time points. These results show that only *react* score and time are significantly related to BDI score across all three time points. Data used in linear models with BDI are shown in Figure 6 (p. 53).

LME modelling was also used to investigate the association between BDI and all non-verbal factor scores over all time points. BDI was the outcome variable, and all non-verbal factor scores, time point, and the interactions between time and non-verbal factor scores were included as fixed effects. Patient ID was included as a random intercept. Model comparison using ANOVA showed that inclusion of the time by *react* score interaction significantly improved the model for BDI ($X^2(2)$= 9.12, p=0.010). The coefficient of this interaction term shows that slope of the regression line between *react* score and BDI decreases by -15.75±4.89 (SE) between pre-op and rough patch and by -11.59±4.27 (SE) between pre-op and 6-months. Model comparison using ANOVA also revealed that *engage/fidget* score significantly improved the LME model for BDI ($X^2(1)= 4.87$, p=0.027), decreasing BDI it by -4.88 ± 2.11(SE) across all three time points. However, the time by *engage/fidget* score interaction was not significant in this model. Finally, time individually decreased BDI by -19.45±3.64 between pre-op and 6-months, and by -16.11±3.81 (SE) between pre-op and rough patch. Thus, this analysis shows that the greatest change in the relationship between BDI score and *react* score occurs in the pre-op to rough patch time interval. This provides information about the timing of behavioral changes related to reactivity during treatment. Additionally, a negative association between BDI and *engage/fidget* score emerges when including the rough patch data in the model. This provides information about the pattern of changes in behaviors related to engagement and fidgeting during treatment.

Association between HDRS-17 and non-verbal factor scores:

Linear regression was used to model the association between HDRS-17 and non-verbal factor score over all three time points. The regression model of HDRS-17 as a function of time, *react* score, and their interaction ($F_{5,30}=27.87$, $p<0.01$), produced a significant time by *react* score interaction in the pre-op to rough patch time interval (β=-4.37, p=0.044). It also produced significant main effect of time in the pre-op between rough patch interval (β=-12.72, p=<0.01) and in the pre-op to 6-month time interval (β=-13.98, p<0.01). The other two regression models of HDRS-17, which were the same but replaced *react* score with *engage/fidget* or *disengage* scores, produced only a significant main effect of time. Data used in linear models with HDRS-17 are shown in Figure 7 (p. 54).

LME modelling was used to investigate the association between HDRS-17 and all non-verbal factor scores over all time points. HDRS-17 was the outcome variable, and all non-verbal factor scores, time point, and the interactions between time and non-verbal factor scores were included as fixed effects. Patient ID was included as a random intercept. Model comparison using ANOVA showed that *engage/fidget* score significantly improved the model ($X^2(1)= 6.57$, p= 0.010), decreasing HDRS-17 by about -1.70±0.63 (SE) across all time points. It also revealed that *react* score significantly improved the model ($X^2(1)= 8.20$, p=0.004), decreasing HDRS-17 score by about -1.91±0.63 (SE). Lastly, it showed that time significantly improved the model ($X^2(1)= 49.58$, $p<0.01$), decreasing HDRS-17 in the pre-op to rough patch interval by -11.63±1.33(SE) and in the pre-op to 6-month interval by -12.38±1.29 (SE). Thus, though these analyses using HDRS-17 were supplementary to the BDI analysis, they produced similar results and therefore provide evidence that verbal rating scores are associated with behaviors related to reactivity, engagement, and fidgeting.

Association between BDI and individual non-verbal behavior scores:

In order to further analyze the influence of individual behaviors on BDI over time, linear regression was used to model the association between BDI score and each behavior score individually over all three time points. Linear regression models were constructed with each behavior, time, and the interaction between behavior score and time as predictors, and BDI as the outcome variable. Four individual behaviors were significantly related to BDI over time. *Pause* increased BDI by 4.76±3.02(SE), *illustrative gestures* decreased BDI by -4.15±1.85(SE), *laugh* decreased BDI by -5.62±1.89(SE), and *smile* decreased BDI by -4.52±1.99(SE). The individual effects of all behavior scores on BDI are shown in Supplementary Figure 1 (see p. 56). In addition, the interactions of time by *head to the side* score and time by *pause* score had large effects on BDI. The time by *head to the side* score decreased BDI by -12.37±4.77(SE) between pre-op and rough patch and by -9.69±4.49 (SE) between pre-op and 6-months. The time by *pause* score interaction decreased BDI by -10.61±1.84 (SE) between pre-op and rough patch, but increased it by 4.84±1.71 (SE) between pre-op and 6-months. Thus this analysis shows that *head to the side* score has a pronounced negative relationship with BDI and that this relationship becomes more extreme over time. Also, *pause* has a pronounced positive relationship with BDI at pre-op and 6-months, but a negative relationship at rough patch. Supplementary Figure 2 (p. 58) shows the distribution over the three time points of three individual behaviors, *pause, illustrative gestures,* and *head to the side,* which are each representative of a non-verbal factor, *disengage, engage/fidget,* and *react,* respectively.

LME modelling was used to investigate the association between BDI and all non-verbal behavior scores over all time points. BDI was the outcome variable, and all behavior scores, time point, and the interactions between time and behavior scores were included as fixed effects. Patient ID was included as a random intercept. The following behaviors improved the LME model compared to a null model using ANOVA: *pause, head to the side, smile, twist mouth, slow speaking, raised eyebrows, no shaking, looking, lip corners in, lip corners back, laugh, illustrative gestures, head bob.* This analysis shows that all these individual behaviors together are important for predicting BDI.

Discussion

This study evaluates the non-verbal behavior of a sample of depressed patients receiving DBS throughout the first 6 months of stimulation. Patients typically experience the greatest improvement during this period, interrupted by a rough patch around 12-14 weeks after the start of stimulation, in which the patients experience an increase in negative feelings/reactivity and typically report symptoms of depression. However, clinicians observe that patients are not acting as severely depressed in their body language as they report through traditional depression rating scales. This study investigates the relationship between non-verbal behavior and traditional depression rating scales, such as the BDI. It also attempts to expand clinical understanding of the rough patch through an analysis of non-verbal behavior using ethnographic methods. This study provides evidence that there is a negative association between non-verbal behaviors related to reactivity and BDI score. This negative association is significantly greater in magnitude at the rough patch and 6-month time points than at it is at the pre-op time point. Additionally, this study shows that there is a significant negative association between BDI score and behaviors related to engagement and fidgeting. Importantly, multiple aspects of these analyses support the conclusion that non-verbal behavior during the rough patch resembles that at 6-months more so than pre-op. This suggests that recovery from depression is already well underway by the time of the 6-month primary clinical outcome time point. These results confirm the importance of considering non-verbal behavior in clinical assessment and validate the impressions of clinicians regarding the well-being of patients during the rough patch.

The ethogram:

Videos of clinical interviews were analyzed using an ethogram, a method used by anthropologists and primatologists alike to record occurrences of discrete non-verbal behaviors. Forty-two behaviors gathered from previous studies were included and deliberately defined in this study's ethogram. After data collection, behaviors that occurred extremely infrequently were excluded from further analysis. This resulted in exclusions of some behaviors that were found to be related to depression severity in other studies. For instance, this study excluded body touching, object touching, and posture shifts, behaviors that were included in Fiquer et al. (2013), Troisi (1999), Ranelli and Miller (1981), and Bouhuys et al. (1991). One potential explanation for the discrepancy between these studies and the present study is that many of our videos included only head and shoulders. Therefore, complete observation of the movements of hands, arms, and legs was not possible in many cases. The lack of observed body-specific movement could also be due to a peculiarity in this sample. As mentioned in the clinical paper by Crowell et al. (2015), this sample of patients is characterized by a lack of reactivity, which could be related to treatment resistance and symptom severity more extreme than patients included in other studies.

Factor analysis:

Factor analysis revealed three groups of correlated behaviors (Figure 4, p. 51). These factors were named *react, engage/fidget,* and *disengage.* The choice of these names was partially based on the content of the factors. The first contained many behaviors that typically occur in response to a statement made by the interviewer, while the third contained behaviors that are typically associated with emotional detachment from interpersonal interaction. The *engage/fidget* factor was named such because it contains some behaviors that are indicative of social engagement, such as *illustrative gestures* and *elaborative speaking*, as well as several body-focused and iterative behaviors. Such iterative behaviors were categorized by Troisi (1999) as "displacements," behavior that "appears in situations characterized by social tension" (Troisi 1999, p. 909). Similarly, Fiquer et al. (2013) defines self-touching "adaptive devices," as behaviors that "appear to help an individual relieve tension" (Fiquer et al. 2013, p. 1117). I chose to characterize such behavior as "fidgets" in order to avoid any *a priori* assumptions about their social meaning. While in EFA, *touch head* was included in the *react* factor, in CFA it was purposely moved to the *engage/fidget* factor in order to align with this conceptual decision. Interestingly, in EFA, *touch head* had a positive factor loading on the *engage/fidget* factor and a negative factor loading on the *react* factor. Conceptually, the results of this study's factor analysis align with factor analyses in other studies of psychomotor behavior in depression. For instance, Ulrich and Harms (1985) found two factors associated with agitation and one associated with retardation when performing factor analysis on the non-verbal behaviors of depressed patients.

Non-verbal behavior factors over time:

The mean frequencies of the *react* and *engage/fidget* factors increased over time, and the mean frequency of *disengage* factor decreased over time. Two way repeated measures ANOVA (ANOVAr) revealed a significant interaction between time and non-verbal factor ($F_{4,44}$= 2.78, p=0.038). This shows that there is a significant difference in the way the scores on the non-verbal behavior factors changed over time. Together with factor analysis, this ANOVA provides evidence that non-verbal behaviors separate into different groups that have different relationships with the recovery process.

A qualitative analysis of Figure 5 (p. 52) can give us more information about how exactly these different factors change over time. Importantly, there is an apparent change in non-verbal behavior that occurs between pre-op and rough patch time points. Specifically, there are high *disengage* scores at pre-op, which decrease by the rough patch time point. The high *disengage* scores indicate higher frequencies of behavior such as slowed speech, which aligns with what has previously been reported about DBS patients at recruitment (Crowell et al. 2015). The high *disengage* scores also indicate lower frequencies of looking at the interviewer, which aligns with other studies that report decreased eye contact in depressed patients (Jones and Pansa, 1979). Another apparent change in behavior between pre-op and rough patch is a slight increase in *react* scores. However, there aren't noticeable changes in the non-verbal factor scores between rough patch and 6-months. This gives us some clues about the process of recovery, indicating that a shift in non-verbal behavioral repertoire, decreased disengagement and increased reactivity, must occur before the rough-patch time point.

This analysis provides evidence that non-verbal behavior is complex and contains various measures that have distinct patterns of variation with improvement. These variations cannot be captured by a single omnibus score of retardation or agitation, as is included in, for instance, the HDRS-17. Thus, non-verbal behavior is a rich and complex source of information, analysis of which can give us a deeper understanding of recovery than what is gained from the sole use of traditional rating scales.

Association between BDI and non-verbal factor scores:

One important question of this study is whether the relationship between non-verbal behavior and verbalized symptom severity is different during the rough patch than at either pre-op or 6-months. If it is, this would indicate that during the rough patch, there is a disconnect between the patients' observable behavior and how they are reporting/interpreting their experience.

The first step in answering this question is to consider if any non-verbal factor scores co-vary with BDI at pre-op and 6-months. The analysis of this first step was performed under the assumption that if associations were found between verbal and non-verbal scores at pre-op and 6-months, these associations are due to the accuracy of the BDI in evaluating depression. In other words, at pre-op and 6-months, BDI is accurate in terms of agreeing with clinicians' intuitive interpretation of non-verbal behavior. In this portion of the analysis, linear regression and LME modelling showed that an increase in behaviors in the *react* factor is associated with a decrease in BDI score between the pre-op and 6-month time points.

After establishing this relationship, the next question is whether the rough patch time point fits into this trend. To answer this question, the same analyses were conducted with all three time points to see if adding the rough patch data would change the resultant model (i.e. change the coefficients of the terms in the regression equation). The model produced with three time points had only minor differences in its coefficients, which indicates that the rough patch time point does follow the trend of the other two. This result does not support the idea that the rough patch identifies a period in which BDI scores are still high but non-verbal behavior is indicative of improvement. In other words, the rough patch does not represent a group of high-BDI and high-*react* scores that is unique from the other two time points. Because the rough patch data did not change the linear models, the following analyses included data from all three time points.

Despite the fact that rough patch data did not change the model, it is important to investigate the models produced with the data from all three time points, and to determine what the coefficients of the terms in the regression equations can tell us about the relationship between BDI and non-verbal behavioral factors. Firstly, there is a negative association between non-verbal behaviors related to reactivity and BDI score. However, one aspect of Figure 6 (p. 53) does not seem to support this result. Rather than having negative slopes between *react* score and BDI at all three time points, the slope at pre-op appears to be positive. However, this positive association is substantially driven by a single outlier with high BDI and very high *react* score. In fact, the correlation between the two scores at pre-op becomes negative when this outlier is removed (changes from a Pearson's R of 0.581 to -0.148). Thus, the coefficients associated with the relationship between *react* score and BDI provide evidence that *react* score increases as BDI decreases over time.

Staying within the *react* factor, the fact that adding the rough patch data did not change the model reveals something interesting about the relationship between *react* factor scores and BDI at this time point. Namely, the *react* scores at rough patch resemble 6 months more than pre-op. The evidence for this conclusion is the coefficient for the time by *react* score interaction term. It is -15.75 between pre-op and rough patch and -11.59 between pre-op and 6-months. This means that the slope of the regression line between *react* score and BDI decreases by -15.75 between pre-op and rough patch and by -11.59 between pre-op and 6-months. Therefore, the relationship between BDI and *react* score changes significantly between pre-op and rough patch. However, after rough patch, there is not a large change; the relationship between BDI and *react* score is similar at rough patch and 6-months. A qualitative look at Figure 6 (p. 53) aligns with this interpretation; pre-op data points create distinct clusters that are noticeably separated from rough patch and 6-month data points. Thus, these results show that *react* score and BDI have a negative association, and that a change in the non-verbal behavior within the *react* factor takes place before the rough patch time point.

Additionally, LME modelling shows that over all three time points, there is a significant negative association between *engage/fidget* score and BDI. Interestingly, this significant negative association emerged in the LME model after including the rough patch data. In the model that only included pre-op and rough patch, the coefficient for *engage/fidget* score is -2.21, while in the model that included all three time points, the coefficient for *engage/fidget* score is -4.88. Thus, the magnitude of the negative association increased when adding rough patch data, which means that the rough patch had to contain many low BDI & high *engage/fidget* data points. This indicates that the majority of the increase in the behaviors in the *engage/fidget* factor over the 6 months occur during the transition between pre-op and rough patch. This aligns with the other evidence suggesting that a unique change in non-verbal behavior occurs between pre-op and rough patch.

In summary, *react* and *engage/fidget* scores are both significantly negatively correlated with BDI. Also, the fact that the rough patch data does not change the regression model reveals an interesting characteristic of this time point: that non-verbal behavior at rough patch resembles that at 6 months. Indeed, the mean BDI score of the rough patch is close to that at 6 months. Also, this analysis aligns with conclusions made from the ANOVA analysis of non-verbal factor scores over time.

HDRS-17 and BDI:

The process of using LME modelling to assess the effect of time and non-verbal factor score on verbal depression score was performed for both BDI and HDRS-17. The results gained from using HDRS-17 as the outcome variable closely resembled those of BDI; they also produced significant main effects of *react* score, *engage/fidget* score, and time. This discussion focuses on interpretation using BDI because the question of interest in this project is the relationship between the patient's conscious experience of their depression and the unconscious physical manifestations of their depression that are observable to others. Since the HDRS-17 is interviewer-rated, it integrates some evaluation of the patient's observable non-verbal behavior. Therefore, it does not solely reflect the patient's experience. On the other hand, BDI is a paper form filled out by the patient; therefore, it is free of the potential confounding of the interviewer's unconscious bias.

The behaviors within the factors:

It is interesting to consider what behaviors constitute the *react* factor, which was consistently found to be related to BDI score and in which changes over time are related to changes in BDI score over time. The individual behavior scores that had the largest correlations with BDI were: *head to the side, pause, illustrative gestures, laugh,* and *smile.* As expected, *head to the side* is a major constituent of the *react* factor, with a factor loading of 0.91. *Laugh* and *smile* are also important constituents of the *react* factor, with loadings of 0.76 and 0.64, respectively. It makes intuitive sense to associate laughing, smiling, and head movement with a decreased BDI score. These details also agree with previous studies of the psychomotor symptoms of depression (Jones and Pansa, 1979; Ranelli and Miller, 1981). Troisi (1999) categorizes *head to the side* as an affiliative gesture, which indicates that its increase with decreasing BDI score could be related to an increase in social connection.

In addition, *pause* score has an interesting relationship with BDI score. *Pause* is part of the *disengage* factor, which was not found to be significantly associated with BDI changes over time. However, individually, *pause* score decreases over time with clinical improvement. Looking within individual time points, *pause* score is positively correlated with BDI score at pre-op and 6-months, which aligns with the previous statement. However, there is an odd reversal of this association at rough patch, in which *pause* score and BDI are negatively correlated. This is reflected in the coefficient for the time by *pause* score interaction term. This coefficient is -10.60 in the pre-op to rough patch time interval, but is 4.84 in the pre-op to 6-months interval. This means that the slope of the regression line between time and *pause* score decreases by -10.60 between pre-op and rough patch, and increases by 4.84 between pre-op and 6-months. This reflects the fact that correlations between *pause* and BDI at pre-op and 6 months are positive (Pearson's R of 0.58 and 0.41, respectively) while at rough patch it is negative (Pearson's R of -0.50). If the relationship between *pause* and BDI followed the same pattern at rough patch as it does at the other two time points, all correlations would be positive. In other words, given the correlations at pre-op and 6-months, people with low *pause* scores at rough patch would be expected to have lower BDIs than what is reported. This indicates that the reported measures don't reflect the non-verbal behavior as accurately at rough patch as they do at the other time points.

This is an important result because the frequency of pausing in speech is a salient measure in many other studies of psychomotor symptoms of depression, as well as being a noticeable characteristic of this population of DBS patients (Crowell et al. 2015). Per Sobin et al.

(1997), speech is a complex motor act in which psychomotor retardation becomes particularly apparent. In line with this interpretation, other studies have shown that the amount of time between utterances is correlated with psychomotor retardation and that depressed patients have increased pause times during counting tasks and clinical interviews compared to healthy individuals (Sobin et al. 1997, Ranelli and Miller 1981).

Lastly, the *illustrative gestures* score was also related to BDI score. This behavior was included in the *engage/fidget* factor. Since this factor was a significant predictor of BDI when analyzing all three time points in regression, it is possible that *illustrative gestures* played a role in driving the significance of this factor. The frequencies of three individual non-verbal behaviors, *pause, illustrative gestures,* and *head to the side*, at the three time points, are shown in Supplementary Figure 2 (p. 58).

Importance for clinical assessment and future directions:

Together, these results emphasize the importance of considering non-verbal behavior in clinical assessment. While body language may be intuitively interpreted by clinicians, ethograms may provide a possible avenue for systematizing and quantifying such interpretations. While this study validates the importance of scores like the BDI and HDRS-17 by showing that many non-verbal behaviors correlate with these scores, it also shows that it is important to consider non-verbal behavior in its own right, especially during transitional phases of recovery. The rough patch is an example of such a phase, when patients are learning new ways of interpreting their experiences. During this time, though patients may have high BDI scores, they display non-verbal behavior that resembles their behavior at 6-months, when their depression rating scores are significantly improved. One possible explanation is that recovery through DBS is reflected in non-verbal behavior sooner than in conscious—and thus verbally communicated—experience. Regardless, this study affirms the utility of paying attention to non-verbal behavior and analyzing it systematically in order to gain the most accurate understanding of recovery.

Future directions could involve validation of the behaviors within the *react* factor as predictors of well-being, ultimately with the goal of identifying the most important non-verbal behaviors that could be attended to in real-time and thus aid in clinical assessment. In addition, these future studies could explore the relationships between these behaviors and other traditional depression and affective scales, such as the Montgomery-Asberg Depression Rating Scale.

A crucial future direction for the analysis of non-verbal behavior in DBS patients is to determine when such a shift in behavior occurs, independently of scores on traditional depression rating scales. Answering this question would involve collecting data on non-verbal behavior at weekly time points between pre-op and rough patch. Determining the timing of the shift in non-verbal behavior would also have import on the topic of prognosis; could an early shift in non-verbal behavior indicate future response or remission? On a related note, it would be interesting to investigate whether the frequencies of non-verbal behaviors at baseline predict a certain response. This would help in patient selection.

An additional potential future use of ethnographic analyses of non-verbal behavior is in regard to parameter adjustment during the rough patch. Observed spikes in depression rating scores might motivate clinicians to adjust stimulation parameters such as electrode contacts. If these scores do not accurately reflect the trajectory of recovery, such adjustments may not be helpful to the patient, and even impede recovery. Analyses of non-verbal behavior can complement traditional rating score to indicate whether patients are actually in need of changes to their stimulation parameters.

Limitations:

This study is relevant for a specific subtype of depressed patients who are receiving DBS and provides insight into their process of recovery. Thus, one limitation of this study is that it assesses only a narrow sample of depressed patients. Due to the fact that psychomotor symptoms may be one of the strongest indicators of the melancholic subtype of depression (Sobin et al. 1997), this may be an important limitation when considering the relevance of these results for other samples of depressed patients. However, the specificity of this study also increases its relevance for this clinical population. Another limitation is that this study does not control for age and sex in its analyses. This could be a serious limitation, as mentioned in Sobin et al. (1997) "both sex and age may be determinants of the manifestation of psychomotor symptoms. Males have been found to have more retardation than females, while females have been found to have more agitation than males" (p. 9). Future studies with larger samples should control for these factors. Another limitation is that non-verbal behavior was only assessed at three physician-defined time points, while a complete analysis would include frequent assessment in order to track any fluctuations in non-verbal behavior over time. Indeed, the choice of rough patch data points here appears to have preceded a larger increase in depression scores occurring at slightly later time points in the study, which may have limited our ability to detect some non-verbal manifestations of the rough patch. Additional limitations include the fact that there was only one observer who analyzed the clinical videos. Though videos were randomized and the observer was blinded to the time point of the video, future studies should investigate variations between observers and consider how this may affect results. A solution to this general problem of observer bias could be the use of automated programs. If a specific set of non-verbal behaviors were determined to have diagnostic or prognostic significance, it is possible that a computer algorithm could be used to count the frequency of such behaviors, eliminating the need for human observers.

Conclusion

The present study demonstrates the utility for analyzing non-verbal behavior in DBS patients throughout the recovery. It shows that non-verbal behavior is a rich source of information about clinical states that cannot be contained by omnibus psychomotor scores on traditional depression rating scores. This study provides evidence that non-verbal behaviors related to reactivity increase as patients become well, particularly the act of tilting the head to the side in what some have interpreted as an affiliative gesture, but which could conservatively be interpreted as a general expressive head movement. This study also provides evidence that several behaviors related to social engagement or disengagement are related to clinical depression scores.

In addition, this study suggests the importance of attention to non-verbal behavior particularly during transitional phases in which patients learn to discriminate transient from chronic negative affect. Because traditional rating scales do not include detailed assessment of non-verbal behavior, the results they generate have not aligned with clinical intuition regarding the body language of patients during such phases, such as the so-called rough patch. This study provides evidence that in DBS patients, non-verbal behavior during the rough patch resembles that at 6-months, when their scores on traditional depression rating scores are significantly improved. Intriguingly, the frequency of pausing during speech at the rough patch implies lower scores on traditional rating scales than what is observed.

This is a pilot study that successfully provides evidence for the importance of assessment of non-verbal behavior in understanding depressive states. An expanded evaluation of ethological methods in a larger sample of DBS patients could provide more conclusive evidence for their utility in diagnostic settings. Results here indicate that this is a promising venture.

Tables and Figures

Table 1. Ethogram

Behavior	*Description*
Eyes	
Looking	**in the direction of the face of the other person**
Look away	**from the interviewer**
Look down	**at feet, lap, or floor**
Flash eyebrows	**quick raising and lowering of eyebrows**
Raised eyebrows	**eyebrows stay up for at least 2 seconds**
Furrowed brow	**wrinkles appear between eyebrows, not an emotional frown**
Head Movement	
Yes nodding	
No shaking	
Head to side	**head is tilted to the side**
Head down	**head is tilted down**
Head up	**head is tilted upward**
Head aversion	**horizontal change in the head position 30 degrees or more to either the left or the right, from the line from interviewee to interviewer**
Bob	**sharp upward movement, like an inverted nod**
Mouth	
Symmetric smile	**smiling in which the muscle that orbits the eye is active in addition to the muscle that pulls the lip corners up; mouth corner retraction occurs in a similar and synchronous way**
Asymmetric smile	**smiling in which the muscle orbiting the eye is not active, mouth corner retraction occurs in an uneven way, showing a "wry" smile**
Lip corners down	**lips are angled down at the corners, usually stretched horizonally**
Tight lips	**lips are pressed together**
Lip corners back	**corners back but not drawn up in a smile**
Lip corners in	**corners of the mouth drawn inward**
Twist mouth	**lips are closed, pushed forward, and twisted to one side**
Lick lips	**tongue passes over lips**
Bite lips	**one lip drawn into the mouth and held there between the teeth**
Expression	
Frown	eyebrows are drawn together, lowered at the center, lip corners down
Cry	**sad face (downturned mouth and lowered brows), including crying**
laugh	**mouth corners drawn up and out, lips part to reveal teeth**
Arm Gestures	
Illustrative gestures	**hand and arm movement used to support speech**
Object using hand movement	purposeful manual activity such as drinking/eating
Shrug	
Body-focused adaptive gestures:	
Light body touching	makes contact with the body; only hand or fingers moving
Intensive body touching	makes contact with the body: forearm is moving
Groom	**hands picking or scratching at hair, face, or body. Includes combing hair**
Touch hair	**hands touch or rub hair on head**
Touch face	**hands touch or rub head or neck**
Touch mouth	**hands makes contact with mouth**
Object-focused adaptive gestures:	
Object touching	plucking or rubbing clothes or other object
Posture	
Lean forward	from the hips towards the interviewer
Posture shift (settle)	shift not forward or backwards; sliding across chair at least 1 inch; slouching movement; curving of the spine; abrupt straightening of the spine
Speech	
Elaborative speaking	**answering the question in more than one sentence**
Slow speaking	**speaking with a <1-1 second pause between each word**
Silence	**3 seconds or more before start of speech or without subsequent speech; not related to listening**
Pause	**2-3 second silence in the middle of two segments of speech**
Verbal backchannel	**affirmative vocalizations produced while listening**

Table 1. The ethogram used to analyze the non-verbal behavior in the videos of patient interviews. Contains behaviors shown in the literature to be related to depression, as well as descriptions of their constituent elements. Bolded items were used in the final data analysis; non-bolded items occurred very infrequently and were excluded from further analysis.

Table 2. Factor scores scaled with mean of zero

		Factor 1	*Factor 2*	*Factor 3*		
Patient ID	**Time Point**	**React**	**Engage/fidget**	**Disengage**	**HDRS-17**	**BDI**
32	pre-op	-0.798	-0.661	-0.518	21	28
33	pre-op	-0.401	-0.488	-0.225	24	46
34	pre-op	-0.911	1.667	-0.545	21	38
35	pre-op	-0.573	-0.594	2.364	26	35
36	pre-op	1.559	-0.656	0.075	23	55
37	pre-op	-0.280	-0.591	-0.520	23	35
901	pre-op	-0.819	-1.012	3.692	26	42
903	pre-op	-0.396	-0.505	-0.526	20	31
905	pre-op	-0.449	-0.589	0.398	24	29
906	pre-op	-0.671	0.639	-0.531	14	21
907	pre-op	-0.271	0.249	1.124	23	32
908	pre-op	-0.569	-0.497	-0.523	24	22
32	rough patch	-0.861	-0.425	-0.510	16	27
33	rough patch	0.172	-0.281	-0.520	14	41
34	rough patch	0.207	2.310	-0.516	5	1
35	rough patch	0.987	0.345	-0.520	6	8
36	rough patch	1.657	0.158	-0.524	4	8
37	rough patch	-0.546	0.005	-0.529	15	28
901	rough patch	0.808	-0.634	-0.546	4	2
903	rough patch	-0.286	-0.397	-0.523	10	28
905	rough patch	0.909	0.008	1.429	5	3
906	rough patch	0.437	0.012	0.075	11	22
907	rough patch	0.597	0.224	-0.520	9	22
908	rough patch	-0.774	0.549	-0.524	7	12
32	6-months	-0.862	-0.678	-0.524	16	26
33	6-months	0.177	0.068	-0.529	12	29
34	6-months	-0.871	3.237	-0.517	7	0
35	6-months	1.349	0.032	0.111	6	3
36	6-months	2.411	-0.571	0.108	6	0
37	6-months	0.882	-0.807	-0.541	6	19
901	6-months	-0.955	-0.624	-0.229	7	5
903	6-months	0.201	0.517	-0.527	8	22
905	6-months	-0.725	-0.339	0.819	15	28
906	6-months	-0.322	0.365	2.353	6	16
907	6-months	1.114	0.444	-0.534	7	25
908	6-months	-1.126	-0.482	-0.530	6	15

Table 2. The model of factor loadings generated through confirmatory factor analysis was used produce factor scores for each patient at each time point, using the original scaled sample. This table shows the resulting factor scores, scaled with a mean of zero.

Figure 1. Location of the electrodes implanted during DBS surgery.

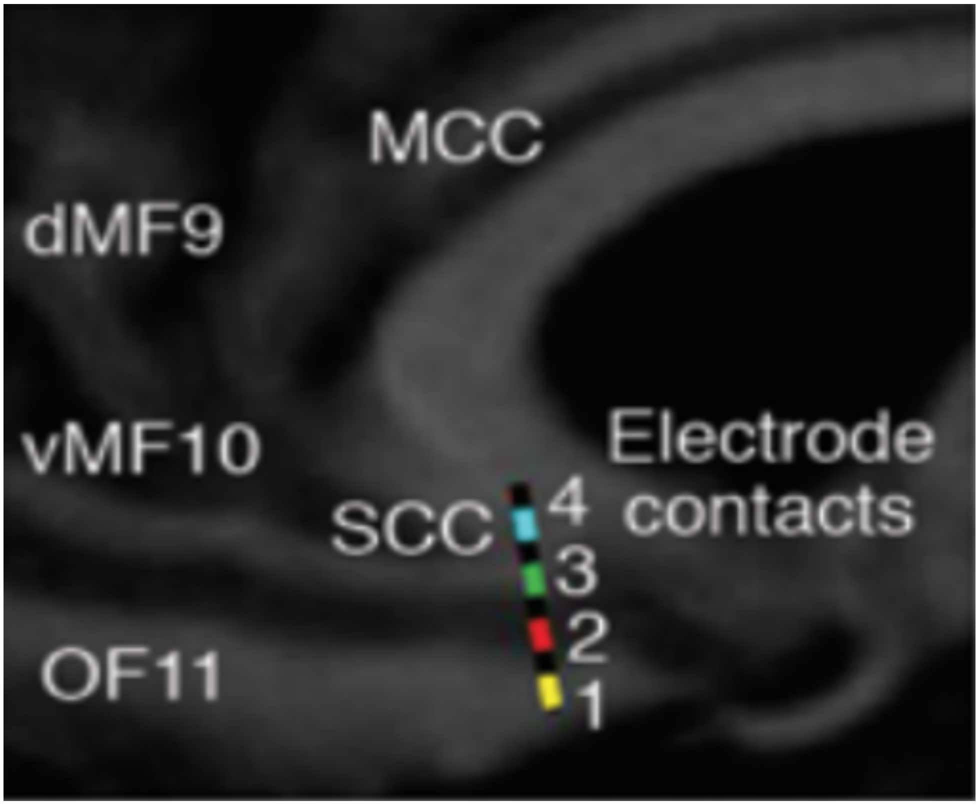


Figure 1. This picture shows the location of the DBS stimulation electrode in the subcallosal cingulate white matter.

Figure 2. Average BDI scores every 2 weeks for the first 6 months of stimulation.

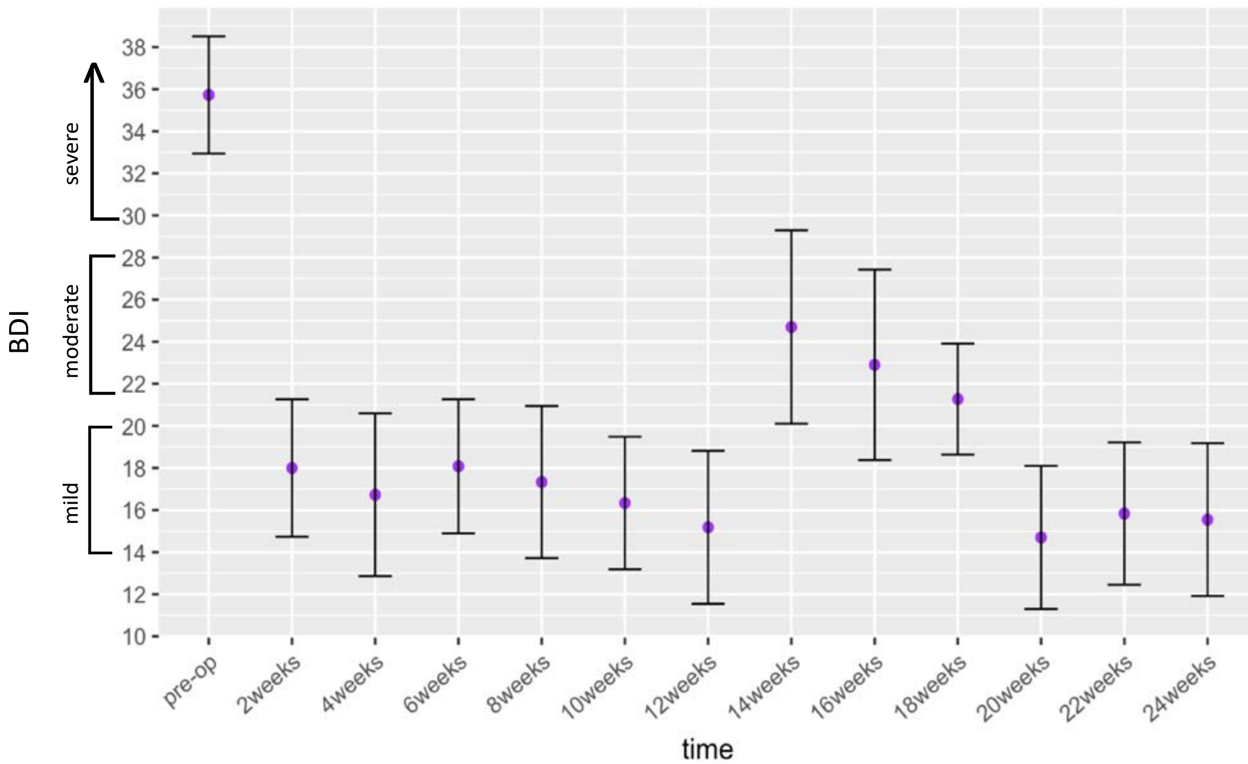


Figure 2. This graph shows the average BDI score of the patients every 2 weeks for the first 6 months of stimulation. Purple dots represent the averages and errorbars represent the standard error of the mean. A BDI score of 14+ is mild depression, 20+ is moderate, and 29+ is severe. The purpose of this graph is convey the non-linear nature of recovery, i.e. the fluctuations in verbal rating score across the first 6 months of stimulation.

Figure 3. Correlation plot of non-verbal behavior scores using Pearson Correlations

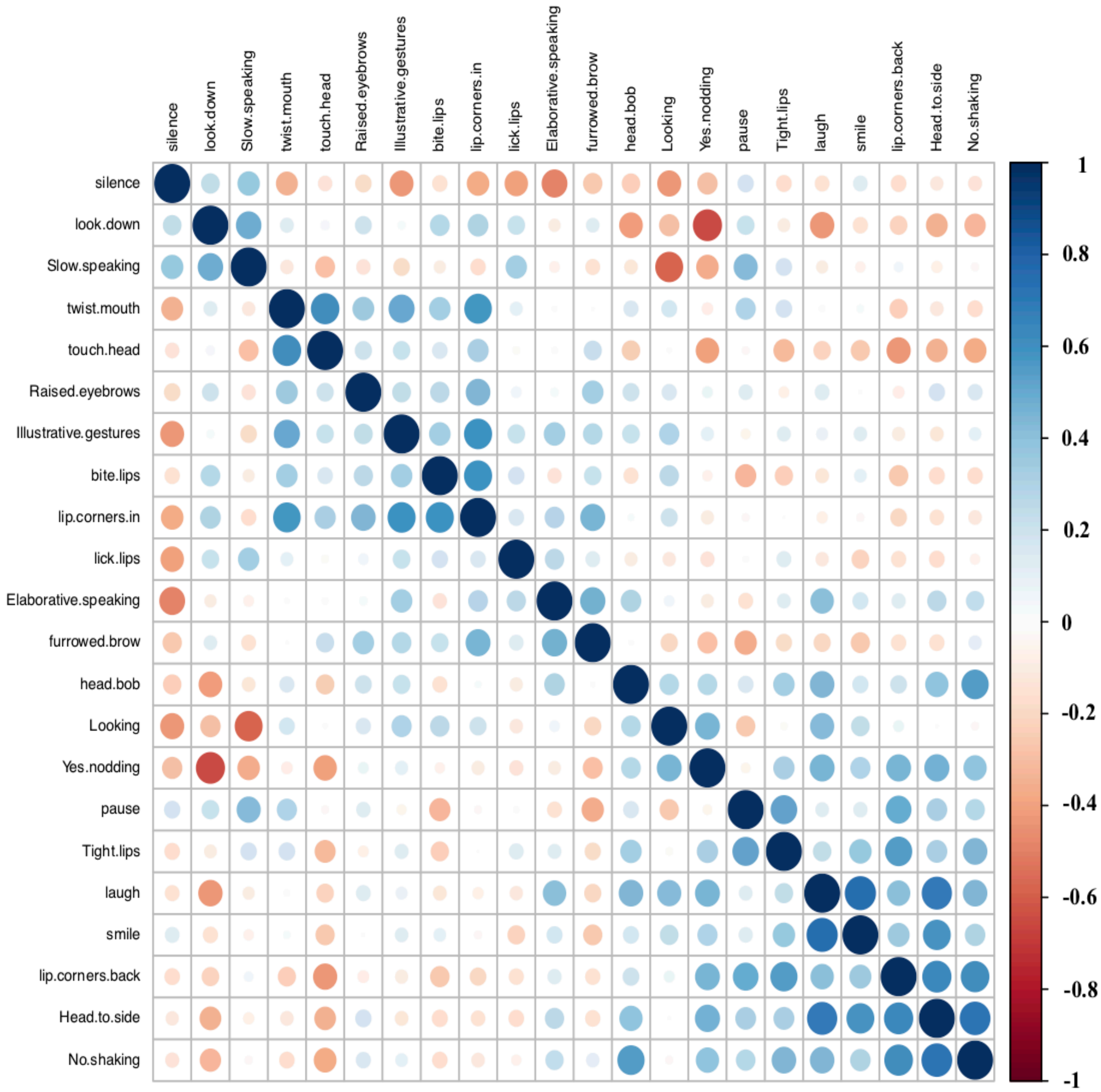


Figure 3. A graphical display of the correlation matrix for all the scores of the non-verbal behaviors included in the final non-verbal factors. These correlations were generated from a bootstrapped sample of 1000 observations for the time points pre-op and 6-months. Color scale corresponding to Pearson's r values are shown in the vertical bar.

Figure 4. Diagram of factor loadings generated from confirmatory factor analysis.

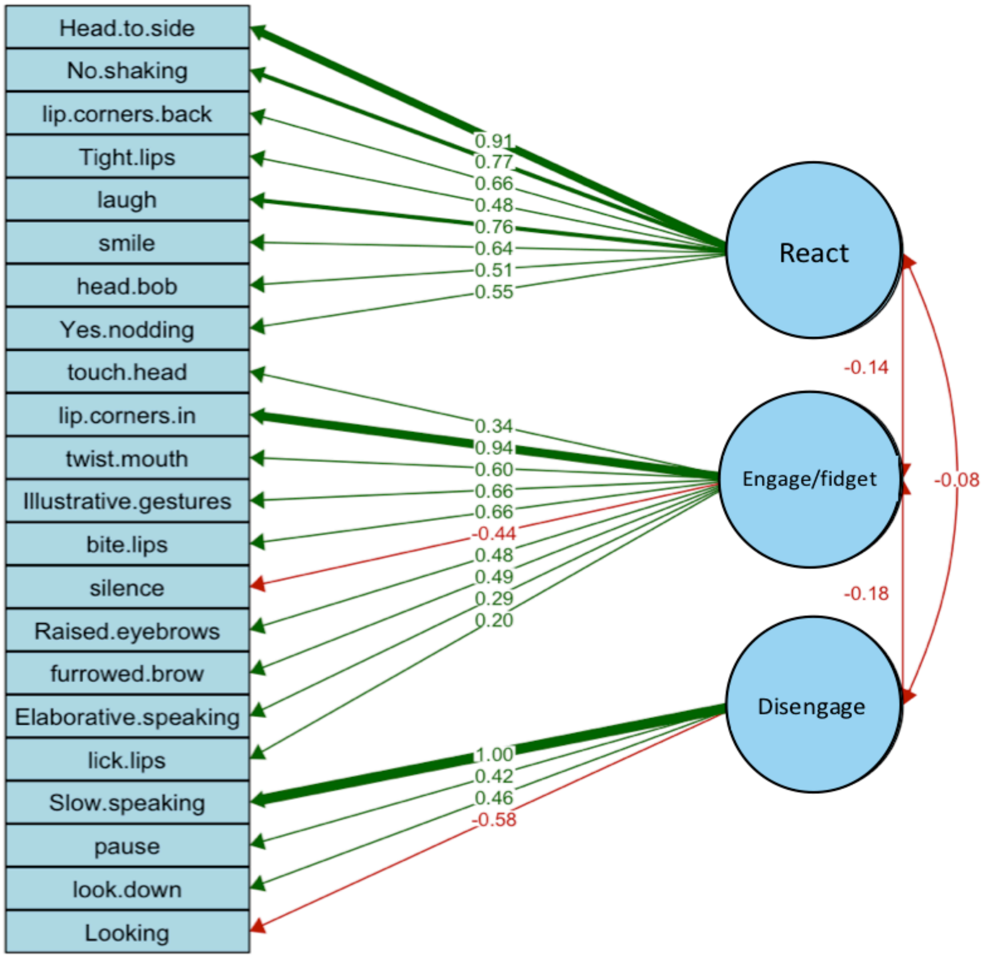


Figure 4. Confirmatory factor analysis (CFA) was performed on a bootstrapped sample of 1000 observations, generated through resampling the original sample. Factor groups were adjusted after exploratory factor analysis to increase conceptual validity. This diagram shows the results of CFA. The three latent factors are inside the circles and the observed items are inside the rectangles. The unidirectional arrows indicate the standardized parameter estimates of the factor loadings of each item on its factor, green indicating positive loading, red indicating negative loading, and width indicating the size of the loading. Bidirectional arrows between factors indicate co-variances.

Figure 5. Non-verbal factor scores over time

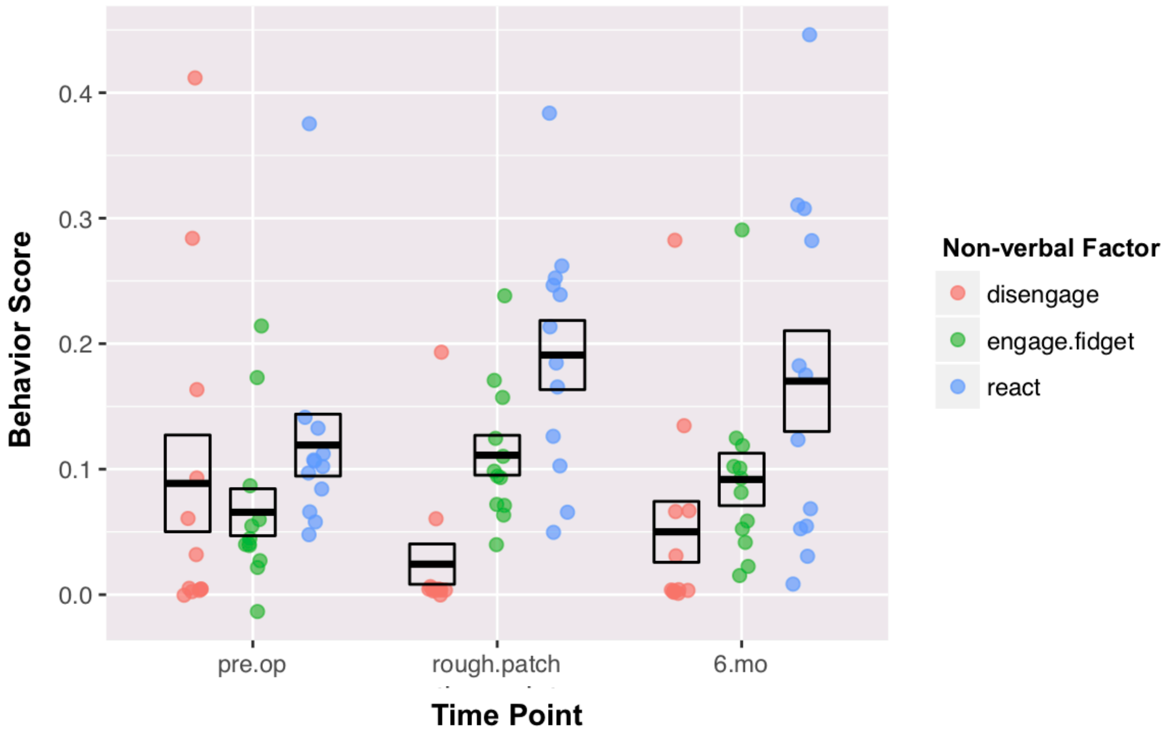


Figure 5. This graph shows the data used in a two-way repeated measures ANOVA. It shows the score of each patient on each factor at each time point. The individual dots are individual patients. Each mean in represented by the boxes' center bars, with the upper and lower bars representing the standard errors of the means.

Figure 6. BDI by non-verbal factor score over time

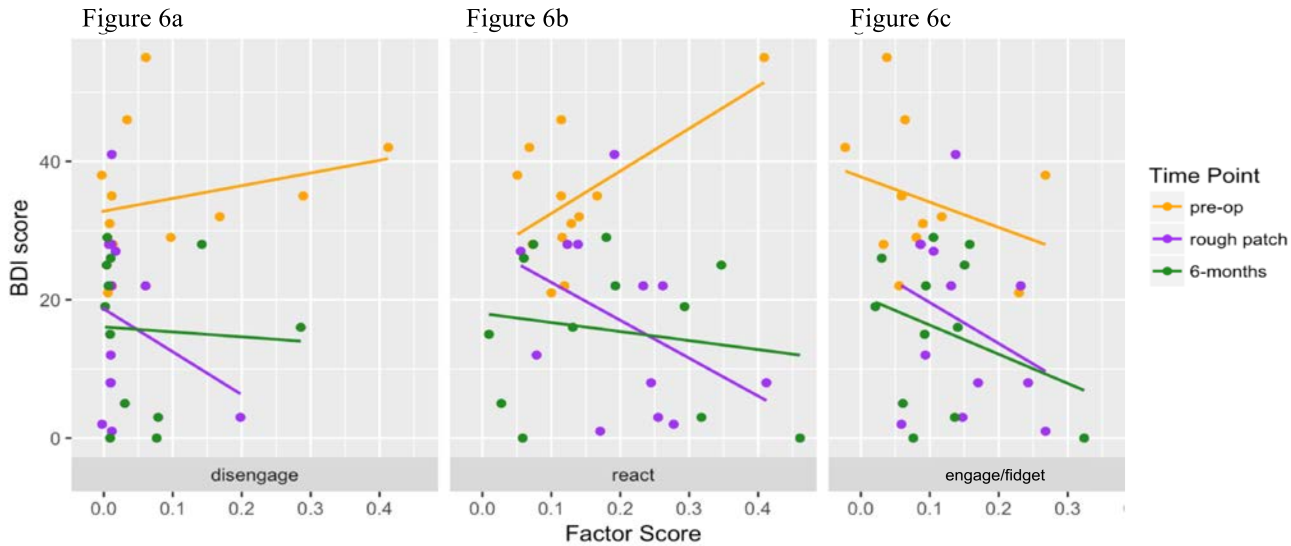


Figure 6. Linear regression was performed for BDI with each scaled factor score individually, time, and their interaction. These graphs show that data used in those analyses, however with original scaling. Regression showed a significant interaction between time and *react* score, as well as a significant main effect of *engage/fidget* score in the prediction of BDI. Each dot is an individual patient. The lines in the graphs represent the linear relationships between the BDI scores and the factor scores.

Figure 7. HDRS-17 by non-verbal factor score over time

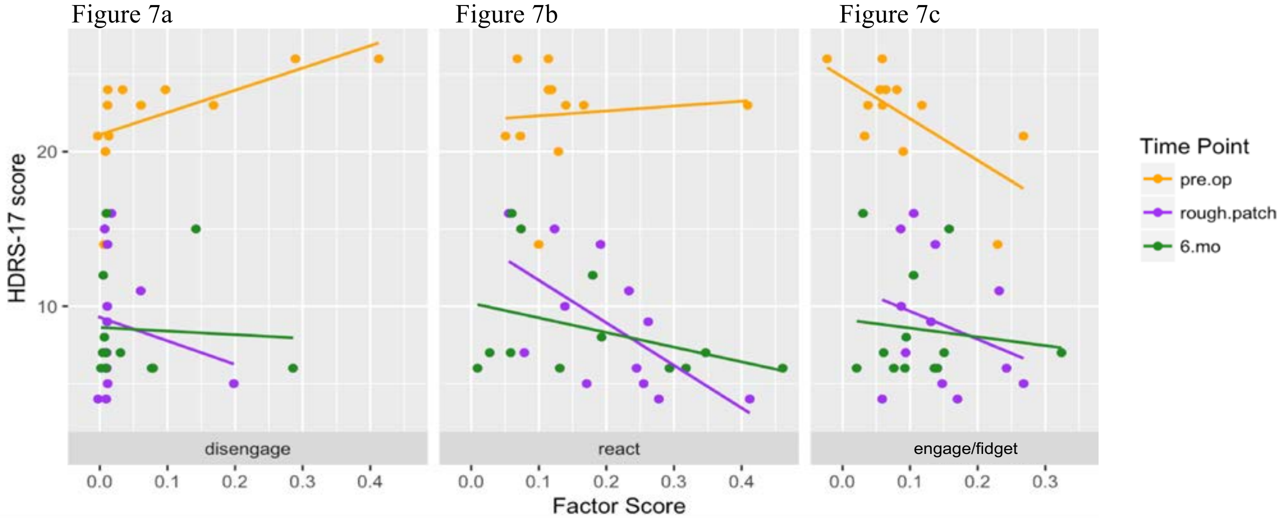


Figure 7. Linear regression was performed for HDRS-17 with each scaled factor score individually, time, and their interaction. These graphs show that data used in those analyses, however with original scaling. Regression showed significant main effects of time and *react* score in the prediction of HDRS-17 score. Each dot is an individual patient. The lines in the graphs represent the linear relationships between the HDRS-17 scores and the factor scores.

Supplementary Table 1. Factor loadings based on exploratory factor analysis

	Factor 1	*Factor 2*	*Factor 2*
Behavior	*React*	*Engage/fidget*	*Disengage*
Head to side	0.828	-0.080	-0.056
No shaking	0.789	0.032	0.053
Lip corners back	0.753	-0.142	0.127
Laugh	0.698	0.027	-0.273
Tight lips	0.682	0.126	0.319
Head bob	0.557	0.163	-0.138
Smile	0.557	0.031	-0.067
Yes nodding	0.536	-0.080	-0.446
Touch head	-0.440	0.286	-0.082
Lip corners in	-0.067	0.908	0.046
Twist mouth	0.025	0.666	0.124
Illustrative gestures	0.080	0.650	-0.099
Bite lips	-0.211	0.537	-0.066
Silence	-0.228	-0.514	0.362
Raised eyebrows	0.113	0.503	-0.022
Furrowed brow	-0.164	0.431	-0.111
Elaborative speaking	0.298	0.325	-0.188
Lick lips	-0.008	0.313	0.184
Slow speaking	0.028	-0.141	0.739
Pause	0.471	0.018	0.676
Look down	-0.293	0.283	0.621
Looking	0.158	0.272	-0.588

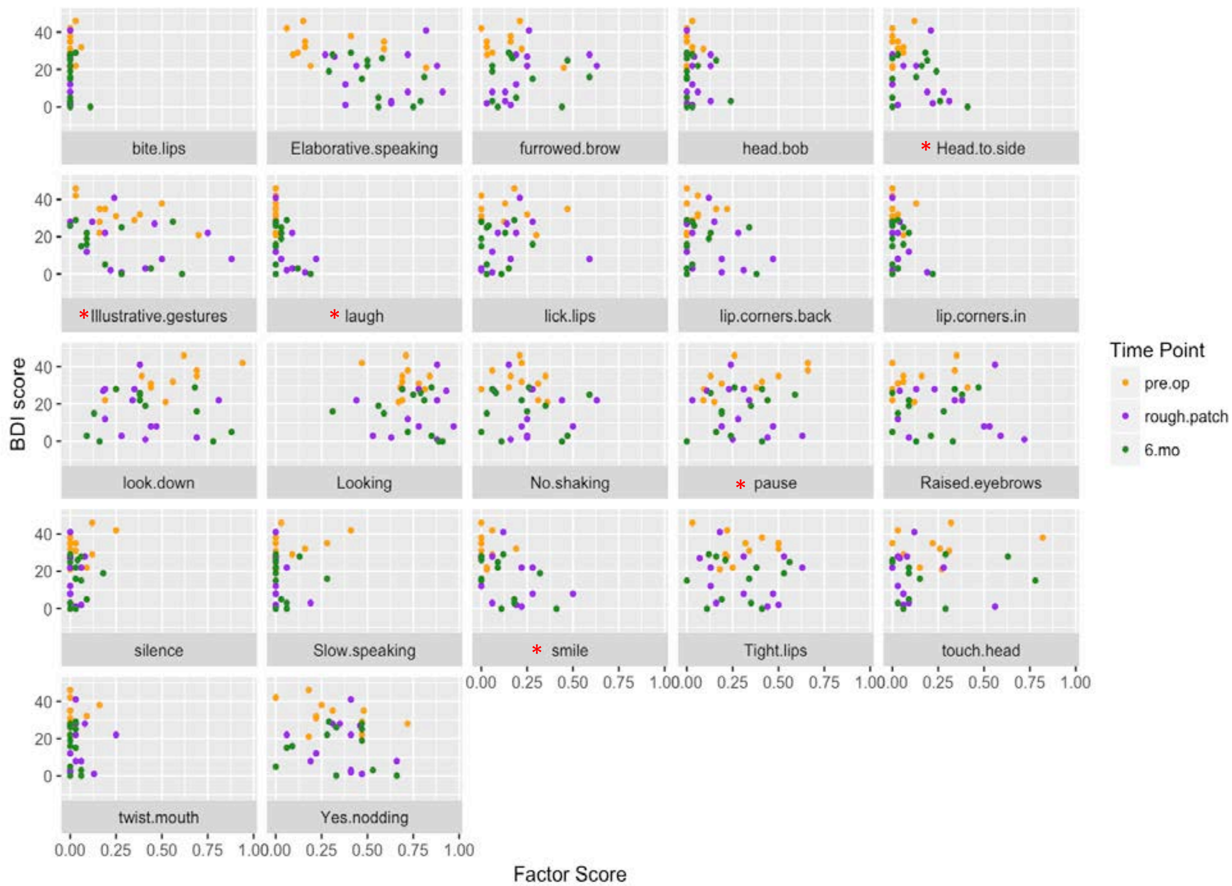
Supplementary Table 1. Exploratory factor analysis (EFA) was performed on a bootstrapped sample of 100 observations, generated through resampling the original sample. This table shows the factor loadings generated through this EFA. Underlined loadings are >0.3 or <-0.3 and indicate the factor to which this item belongs, based on the EFA.

Supplementary Table 2. P-values of pairwise comparisons using paired t-tests

	6 months-disengage	pre op-disengage	rough patch-disengage	6 months-engage/fidget	pre op-engage/fidget	rough patch-engage/fidget	6 months-react	pre op-react
pre op-disengage	0.374							
rough patch-disengage	0.429	0.131						
6 months-engage/fidget	0.200	0.944	0.035					
pre op-engage/fidget	0.554	0.651	0.169	0.396				
rough patch-engage/fidget	0.096	0.579	**0.002**	0.554	0.080			
6 months-react	0.029	0.125	0.010	0.133	0.062	0.170		
pre op-react	0.082	0.540	**0.000**	0.403	0.151	0.800	0.357	
rough patch-react	**0.001**	0.015	**0.000**	0.023	**0.002**	0.018	0.667	0.056

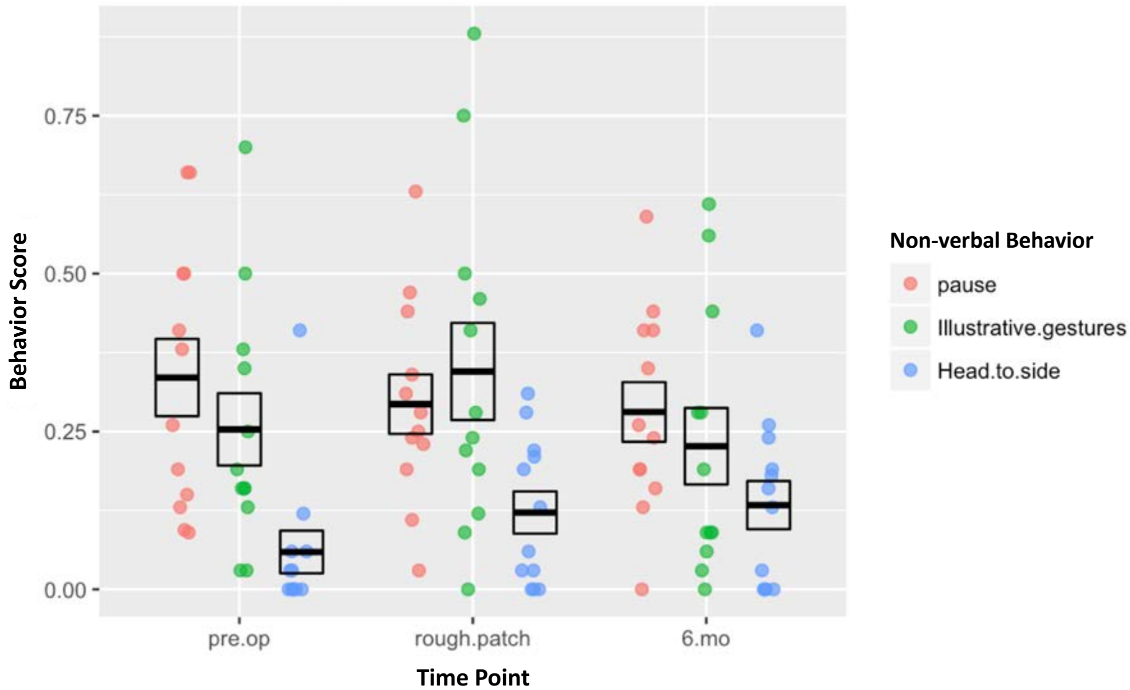
Supplementary Table 2. After two-way repeated measures ANOVA, pairwise comparisons between each factor score at each time point were performed using paired t-tests. The comparisons of interest in this post-hoc analysis were between different time points within each factor, to look at how each factor score changes over time. This table shows the p-values of each t-test. Significance level was adjusted for 6 comparisons of interest (each factor score at rough patch and 6-months compared to that factor score at pre-op) per the following formula: 0.05/6=0.0083. The bolded p-values are significant at the adjusted significance level.

Supplementary Figure 1. BDI by individual behavior scores over time



Supplementary Figure 1. Linear regression was peformed for BDI score with each scaled behavior score in order to evaluate the effect of each behavior score on BDI. These graphs show the relationship between BDI and each behavior with its original scaling score over time. Red stars indicate extreme effects.

Supplementary Figure 2. Pause, illustrative gesture, and head to the side scores over time



Supplementary Figure 2. This graph shows the scores of each patient on three individual examples of non-verbal behaviors at each time point. Each behavior is part of a different factor: *pause* is part of *disengage*, *illustrative gestures* is part of engage/fidget, and *head to the side* is part of *react.* The individual dots are individual patients. Each mean is represented by the boxes' center bars, with the upper and lower bars representing the standard errors of the means.

References

Aberg-Wistedt, A. (1989). The antidepressant effects of 5-HT uptake inhibitors. In *British Journal of Psychiatry* (Vol. 155, pp. 32–40).

Akil, H., Gordon, J., Hen, R., Javitch, J., Mayberg, H., McEwen, B., … Nestler, E. J. (2018). Treatment resistant depression: A multi-scale, systems biology approach. *Neuroscience and Biobehavioral Reviews*, *84*(July 2017), 272–288. http://doi.org/10.1016/j.neubiorev.2017.08.019

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale : Has the Gold Standard Become a Lead Weight ? Reviews and Overviews The Hamilton Depression Rating Scale : Has the Gold Standard Become a Lead Weight ? *Am J Psychiatry*, *161*(December), 2163–2177. http://doi.org/10.1176/appi.ajp.161.12.2163

Barrett, K. (2017). Psychiatric neurosurgery in the 21st century: overview and the growth of deep brain stimulation. *BJPsych Bulletin*, *41*(5), 281–286. http://doi.org/10.1192/pb.bp.116.055772

Benabid, A. L. (2003). Deep brain stimulation for Parkinson's disease. *Current Opinion in Neurobiology*, *13*(6), 696–706. http://doi.org/10.1016/j.conb.2003.11.001

Berlim, M. T., Van Den Eynde, F., Tovar-Perdomo, S., & Daskalakis, Z. J. (2014). Response, remission and drop-out rates following high-frequency repetitive transcranial magnetic stimulation (rTMS) for treating major depression: A systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. *Psychological Medicine*, *44*(2), 225–239. http://doi.org/10.1017/S0033291713000512

Bernaards, Coen A. and Jennrich, Robert I. (2005) Gradient Projection Algorithms and Software for ArbitraryRotation Criteria in Factor Analysis, Educational and Psychological Measurement: 65, 676-696. <http://www.stat.ucla.edu/research/gpa>

Blair, Graeme, Jasper Cooper, Alexander Coppock, Macartan Humphreys and Aaron Rudkin (2018). fabricatr: Imagine Your Data Before You Collect It. R package version 0.2.0. https://CRAN.R-project.org/package=fabricatr

Bouhuys, A. L., Jansen, C. J., & van den Hoofdakker, R. H. (1991). Analysis of observed behaviors displayed by depressed patients during a clinical interview: relationships between behavioral factors and clinical concepts of activation. *Journal of Affective Disorders*, *21*(2), 79–88. http://doi.org/10.1016/0165-0327(91)90053-U

Bouhuys, A. L., & van den Hoofdakker, R. H. (1991). The interrelatedness of observed behavior of depressed patients and of a psychiatrist: an ethological study on mutual influence. *Journal of Affective Disorders*, *23*(2), 63–74. http://doi.org/10.1016/0165-0327(91)90093-8

Bunney, W. E., & Davis, J. M. (1965). Norepinephrine in Depressive Reactions. *Archives of General Psychiatry*, *13*(6), 483. http://doi.org/10.1001/archpsyc.1965.01730060001001

Carlsson, A., Lindqvist, M., & Magnusson, T. (1957). 3,4-Dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists [16]. *Nature*. http://doi.org/10.1038/1801200a0

Chi, K. F., Korgaonkar, M., & Grieve, S. M. (2015). Imaging predictors of remission to anti-depressant medications in major depressive disorder. *Journal of Affective Disorders*, *186*, 134–144. http://doi.org/10.1016/j.jad.2015.07.002

Crowell, A. L., Garlow, S. J., Riva-posse, P., Mayberg, H. S., & Christen, M. (2015). Characterizing the therapeutic response to deep brain stimulation for treatment-resistant depression: a single center long-term perspective. *Frontiers in Integrative Neuroscience*, *9*(June), 0–6. http://doi.org/10.3389/fnint.2015.00041

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Dunlop, B. W., Rajendra, J. K., Craighead, W. E., Kelley, M. E., McGrath, C. L., Choi, K. S., … Mayberg, H. S. (2017). Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. *American Journal of Psychiatry*, *174*(6), 533–545. http://doi.org/10.1176/appi.ajp.2016.16050518

Eitan, R., Fontaine, D., Benoît, M., Giordana, C., Darmon, N., Israel, Z., … Lerer, B. (2018). One year double blind study of high vs low frequency subcallosal cingulate stimulation for depression. *Journal of Psychiatric Research*, *96*, 124–134. http://doi.org/10.1016/j.jpsychires.2017.09.026

Fiquer, J. T., Boggio, P. S., & Gorenstein, C. (2013). Talking bodies: Nonverbal behavior in the assessment of depression severity. *Journal of Affective Disorders*, *150*(3), 1114–1119. http://doi.org/10.1016/j.jad.2013.05.002

Fiquer, J. T., Moreno, R. A., Canales, J. Z., Cavalcanti, A., & Gorenstein, C. (2017). Is nonverbal behavior in patients and interviewers relevant to the assessment of depression and its recovery? A study with Dutch and Brazilian patients. *Psychiatry Research*, *250*(January), 59–64. http://doi.org/10.1016/j.psychres.2017.01.053

Fossi, L., Faravelli, C., & Paoli, M. (1984). The Ethological Approach to the Assessment of Depressive Disorders. *The Journal of Nervous and Mental Disease*, *172*(6), 332–341.

Gaebel, W., & Wölwer, W. (2004). Facial expressivity in the course of schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience*, *254*(5), 335–342. http://doi.org/10.1007/s00406-004-0510-5

Geerts, E., & Brüne, M. (2009). Ethological approaches to psychiatric disorders: focus on depression and schizophrenia. *Australian and New Zealand Journal of Psychiatry*, *43*(11), 1007–1015. http://doi.org/10.3109/00048670903270498

George, M. S., Nahas, Z., Kozel, F. A., Li, X., Denslow, Stewart Yamanaka, K., Mishory, A., … Bohning, D. E. (2002). Mechanisms and State of the Art of Transcranial Magnetic Stimulation. *The Journal of ECT*, *18*(4), 170–181. http://doi.org/10.1097/00124509-200212000-00002

Guinjoan, S. M., Mayberg, H. S., Costanzo, E. Y., Fahrer, R. D., Tenca, E., Antico, J., … Nemeroff, C. B. (2010). Asymmetrical contribution of brain structures to treatment-resistant depression as illustrated by effects of right subgenual cingulum stimulation. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *22*(3), 265–77. http://doi.org/10.1176/jnp.2010.22.3.265

Guo, Q., Li, C., & Wang, J. (2017). Updated Review on the Clinical Use of Repetitive Transcranial Magnetic Stimulation in Psychiatric Disorders. *Neuroscience Bulletin*, *33*(6), 1–10. http://doi.org/10.1007/s12264-017-0185-3

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62. http://doi.org/10.1136/jnnp.23.1.56

Härter, M., Klesse, C., Bermejo, I., Schneider, F., & Berger, M. (2010). Unipolar depression:

diagnostic and therapeutic recommendations from the current S3/National Clinical Practice Guideline. *Deutsches Ärzteblatt International*, *107*(40), 700–708. http://doi.org/10.3238/arztebl.2010.0700

Hilimire, M. R., Mayberg, H. S., Holtzheimer, P. E., Broadway, J. M., Parks, N. A., DeVylder, J. E., & Corballis, P. M. (2015). Effects of subcallosal cingulate deep brain stimulation on negative self-bias in patients with treatment-resistant depression. *Brain Stimulation*, *8*(2), 185–191. http://doi.org/10.1016/j.brs.2014.11.010

Hirschfeld, R. M. A. (2000). History and evolution of the monoamine hypothesis of depression. *Journal of Clinical Psychiatry*, *61*(SUPPL. 6), 4–6. http://doi.org/10.4088/JCP.v61n0201

Holtzheimer, P. E., Husain, M. M., Lisanby, S. H., Taylor, S. F., Whitworth, L. A., McClintock, S., … Mayberg, H. S. (2017). Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. *The Lancet Psychiatry*, *4*(11), 839–849. http://doi.org/10.1016/S2215-0366(17)30371-1

Holtzheimer, P. E., Kelley, M. E., Gross, R. E., Filkowski, M. M., Garlow, S. J., Barrocas, A., … Mayberg, H. S. (2012). Subcallosal cingulate deep brain stimulation for treatment-resistant unipolar and bipolar depression. *Archives of General Psychiatry*, *69*(2), 150–158. http://doi.org/10.1001/archgenpsychiatry.2011.1456

Holtzheimer, P. E., & Mayberg, H. S. (2011). Stuck in a rut: Rethinking depression and its treatment. *Trends in Neurosciences*. http://doi.org/10.1016/j.tins.2010.10.004

Husain, M. M., Rush, A. J., Fink, M., Knapp, R., Petrides, G., Rummans, T., … Kellner, C. H. (2004). Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): A consortium for research in ECT (CORE) report. *Journal of Clinical Psychiatry*, *65*(4), 485–491. http://doi.org/10.4088/JCP.v65n0406

Kedzior, K. K., Reitz, S. K., Azorina, V., & Loo, C. (2015). Durability of the Antidepressant Effect of the High-Frequency Repetitive Transcranial Magnetic Stimulation (rTMS) in the Absence of Maintenance Treatment in Major Depression: A Systematic Review and Meta-Analysis of 16 Double-blind, Randomized, Sham-contr. *Depression and Anxiety*, *32*(3), 193–203. http://doi.org/10.1002/da.22339

Kellner, C. H., Knapp, R. G., Petrides, G., Rummans, T. A., Husain, M. M., Rasmussen, K., … Fink, M. (2006). Continuation Electroconvulsive Therapy vs Pharmacotherapy for Relapse Prevention in Major Depression. *Archives of General Psychiatry*, *63*(12). http://doi.org/10.1001/archpsyc.63.12.1337

Kempton, M. J. (2011). Structural Neuroimaging Studies in Major Depressive Disorder. *Archives of General Psychiatry*, *68*(7), 675. http://doi.org/10.1001/archgenpsychiatry.2011.60

Kennedy, S. H., Giacobbe, P., Rizvi, S. J., Placenza, F. M., Nishikawa, Y., Mayberg, H. S., & Lozano, A. M. (2011). Deep brain stimulation for treatment-resistant depression: follow-up after 3 to 6 years. *American Journal of Psychiatry*, *168*(1535–7228 (Electronic)), 502–510. http://doi.org/DOI 10.1176/appi.ajp.2010.10081187

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama*, *289*(23), 3095–105. http://doi.org/10.1097/00132578-200310000-00002

Lane, R. D., Weidenbacher, H., Smith, R., Fort, C., Thayer, J. F., & Allen, J. J. B. (2013). Subgenual anterior cingulate cortex activity covariation with cardiac vagal control is altered in depression. *Journal of Affective Disorders*, *150*(2), 565–570. http://doi.org/10.1016/j.jad.2013.02.005

Lawrence, Michael A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Lemke, M., Puhl, P., Koethe, N., & Winkler, T. (1999). Psychomotor retardation and anhedonia in depression. *Acta Psychiatrica Scandinavica*, *99*, 252–256.

Lozano, A. M., Giacobbe, P., Hamani, C., Rizvi, S. J., Kennedy, S. H., Kolivakis, T. T., … Mayberg, H. S. (2012). A multicenter pilot study of subcallosal cingulate area deep brain stimulation for treatment-resistant depression. *Journal of Neurosurgery*, *116*(2), 315–322. http://doi.org/10.3171/2011.10.JNS102122

Lozano, A. M., Mayberg, H. S., Giacobbe, P., Hamani, C., Craddock, R. C., & Kennedy, S. H. (2008). Subcallosal Cingulate Gyrus Deep Brain Stimulation for Treatment-Resistant Depression. *Biological Psychiatry*, *64*(6), 461–467. http://doi.org/10.1016/j.biopsych.2008.05.034

Mayberg, H., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., … Objective: (1999). Reciprocal limbic-cortical function and negative mood : Converging PET ... *Am J Psychiatry*, *156*(May), 675–682.

Mayberg, H. S. (2009). Targeted electode-based modulation of neural circuits for depression. *Journal of Clinical Investigation*, *119*(4), 717–725. http://doi.org/10.1172/JCI38454

Mayberg, H. S., Brannan, S. K., Tekell, J. L., Silva, J. a, Mahurin, R. K., McGinnis, S., & Jerabek, P. a. (2000). Regional metabolic effects of fluoxetine in major depression: serial changes and relationship to clinical response. *Biological Psychiatry*, *48*(8), 830–43. http://doi.org/10.1016/S0006-3223(00)01036-2

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., … Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, *45*(5), 651–660. http://doi.org/10.1016/j.neuron.2005.02.014

McGrath, C. L., Kelley, M. E., Holtzheimer ,III, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., … Mayberg, H. S. (2013). Toward a Neuroimaging Treatment Selection Biomarker for Major Depressive Disorder. *Jama Psychiatry*, *70*(8), 821–829. http://doi.org/10.1001/jamapsychiatry.2013.143

Merkl, A., Aust, S., Schneider, G.-H., Visser-Vandewalle, V., Horn, A., Kühn, A. A., … Bajbouj, M. (2017). Deep Brain Stimulation of the Subcallosal Cingulate Gyrus in Patients with Treatment-Resistant Depression: a double-blinded randomized controlled study and long-term follow-up in eight patients. *Journal of Affective Disorders*. http://doi.org/10.1016/j.jad.2017.11.024

Merkl, A., Schneider, G.-H. H., Schonecker, T., Aust, S., Kuhl, K.-P., Kupsch, A., … Bajbouj, M. (2013). Antidepressant effects after short-term and chronic stimulation of the subgenual cingulate gyrus in treatment-resistant depression. *Experimental Neurology*, *249*, 160–168. http://doi.org/10.1016/j.expneurol.2013.08.017

Morishita, T., Fayad, S. M., Higuchi, M., Nestor, K. A., & Foote, K. D. (2014). Deep Brain Stimulation for Treatment-resistant Depression: Systematic Review of Clinical Outcomes. *Neurotherapeutics*, *11*(3), 475–484. http://doi.org/10.1007/s13311-014-0282-1

Moro, E., Lozano, A. M., Pollak, P., Agid, Y., Rehncrona, S., Volkmann, J., … Lang, A. E. (2010). Long-term results of a multicenter study on subthalamic and pallidal stimulation in Parkinson’s disease. *Movement Disorders*, *25*(5), 578–586. http://doi.org/10.1002/mds.22735

Musil, R., Seemüller, F., Meyer, S., Spellmann, I., Adli, M., Bauer, M., … Riedel, M. (2017). Subtypes of depression and their overlap in a naturalistic inpatient sample of major depressive disorder. *International Journal of Methods in Psychiatric Research*, (April 2017), 1–16. http://doi.org/10.1002/mpr.1569

Naesström, M., Blomstedt, P., & Bodlund, O. (2016). A systematic review of psychiatric indications for deep brain stimulation, with focus on major depressive and obsessive-compulsive disorder. *Nordic Journal of Psychiatry*, *70*(7), 483–491. http://doi.org/10.3109/08039488.2016.1162846

Narang, P., Retzlaff, A., Brar, K., & Lippmann, S. (2016). Deep Brain Stimulation for Treatment-Refractory Depression. *South. Med. J.*, *109*(11), 700–703. http://doi.org/10.14423/SMJ.0000000000000554

Neumaier, F., Paterno, M., Alpdogan, S., Tevoufouet, E. E., Schneider, T., Hescheler, J., & Albanna, W. (2017). Surgical Approaches in Psychiatry: A Survey of the World Literature on Psychosurgery. *World Neurosurgery*, *97*, 603–634.e8. http://doi.org/10.1016/j.wneu.2016.10.008

Pedersen, J., Schelde, J. T. M., Hannibal, E., Behnke, K., Nielsen, B. M., & Hertz, M. (1988). An ethological description of depression. *Acta Psychiatrica Scandinavica*, *78*(3), 320–330. http://doi.org/10.1111/j.1600-0447.1988.tb06343.x

Puigdemont, D., Pérez-Egea, R., Portella, M. J., Molet, J., de Diego-Adeliño, J., Gironell, A., … Pérez, V. (2012). Deep brain stimulation of the subcallosal cingulate gyrus: further evidence in treatment-resistant major depression. *The International Journal of Neuropsychopharmacology / Official Scientific Journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, *15*(1), 121–33. http://doi.org/10.1017/S1461145711001088

Ramasubbu, R., Anderson, S., Haffenden, A., Chavda, S., & Kiss, Z. H. T. (2013). Double-blind optimization of subcallosal cingulate deep brain stimulation for treatment-resistant depression: a pilot study. *Journal of Psychiatry & Neuroscience : JPN*, *38*(5), 325–32. http://doi.org/10.1503/jpn.120160

Raiche, G. (2010). nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test.R package version 2.3.3.

Ranelli, C. J., & Miller, R. E. (1981). Behavioral Predictors of Amitriptyline Response in Depression. *American Journal of Psychiatry*, *138*(1), 30–34.

Ren, J., Li, H., Palaniyappan, L., Liu, H., Wang, J., Li, C., & Rossini, P. M. (2014). Repetitive transcranial magnetic stimulation versus electroconvulsive therapy for major depression: a systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *51*, 181–9. http://doi.org/10.1016/j.pnpbp.2014.02.004

Ressler, K. J., & Mayberg, H. S. (2007). Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nature Neuroscience*, *10*(9), 1116–24. http://doi.org/10.1038/nn1944

Revelle, W. (2017) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.7.8.

Rizvi, S. J., Donovan, M., Giacobbe, P., Placenza, F., Rotzinger, S., & Kennedy, S. H. (2011). Neurostimulation therapies for treatment resistant depression: A focus on vagus nerve stimulation and deep brain stimulation. *International Review of Psychiatry*, *23*(5), 424–436. http://doi.org/10.3109/09540261.2011.630993

Rosa, M. A., & Lisanby, S. H. (2012). Somatic treatments for mood disorders. *Neuropsychopharmacology*. http://doi.org/10.1038/npp.2011.225

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *American Journal of Psychiatry*, *163*(11), 1905–1917. http://doi.org/10.1176/appi.ajp.163.11.1905

Sachdev, P., & Aniss, A. M. (1994). Slowness of movement in melancholic depression. *Biological Psychiatry*, *35*(4), 253–262. http://doi.org/10.1016/0006-3223(94)91256-4

Sacher, J., Neumann, J., Fünfstück, T., Soliman, A., Villringer, A., & Schroeter, M. L. (2012). Mapping the depressed brain: A meta-analysis of structural and functional alterations in major depressive disorder. *Journal of Affective Disorders*, *140*(2), 142–148. http://doi.org/10.1016/j.jad.2011.08.001

Schelde, J. T. M., Pedersen, J., Hannibal, E., Behnke, K., Nielsen, B. M., & Hertz, M. (1988). An ethological analysis of depression: comparison between ethological recording and Hamilton rating of five endogenously depressed patients. *Acta Psychiatrica Scandinavica*, *78*(3), 331–340. http://doi.org/10.1111/j.1600-0447.1988.tb06344.x

Schildkraut, J. J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *The American Journal of Psychiatry*. http://doi.org/10.1176/ajp.122.5.509

Siegle, G. J., Thompson, W. K., Collier, A., Berman, S. R., Feldmiller, J., Thase, M. E., & Friedman, E. S. (2012). Toward Clinically Useful Neuroimaging in Depression Treatment. *Archives of General Psychiatry*, *69*(9), 913. http://doi.org/10.1001/archgenpsychiatry.2012.65

Sobin, C., Ph, D., Sackeim, H. A., & Ph, D. (1997). Psychomotor Symptoms of Depression, (January), 4–17.

Speer, A. M., Benson, B. E., Kimbrell, T. K., Wassermann, E. M., Willis, M. W., Herscovitch, P., & Post, R. M. (2009). Opposite effects of high and low frequency rTMS on mood in depressed patients: Relationship to baseline cerebral activity on PET. *Journal of Affective Disorders*, *115*(3), 386–394. http://doi.org/10.1016/j.jad.2008.10.006

Speer, a M., Kimbrell, T. a, Wassermann, E. M., D Repella, J., Willis, M. W., Herscovitch, P., & Post, R. M. (2000). Opposite effects of high and low frequency rTMS on regional brain activity in depressed patients. *Biological Psychiatry*, *48*(12), 1133–41. http://doi.org/10.1016/S0006-3223(00)01065-9

Troisi, A. (1999). Ethological research in clinical psychiatry: The study of nonverbal behavior during interviews. *Neuroscience and Biobehavioral Reviews*, *23*(7), 905–913. http://doi.org/10.1016/S0149-7634(99)00024-X

Vasavada, M. M., Leaver, A. M., Njau, S., Joshi, S. H., Ercoli, L., Hellemann, G., … Espinoza, R. (2017). Short- and Long-term Cognitive Outcomes in Patients With Major Depression Treated With Electroconvulsive Therapy. *The Journal of ECT*, *33*(4), 1. http://doi.org/10.1097/YCT.0000000000000426

Wang, Y.-P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Revista Brasileira de Psiquiatria*, *35*, 416–431. http://doi.org/10.1590/1516-4446-2012-1048

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.